DOCUMENT RESUME

ED 041 896 24 TE 001 948

AUTHOR	King, F. J.; And Others.
TITLE	Effectiveness of Two Ways of Teaching Grammar to Students of Different Ability Patterns. Final Report.
INSTITUTION	Florida State Univ., Tallahassee.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C. National Center for Educational Research and Development.
BUREAU NO	BR-8-D-023
PUB DATE	Oct 69
GRANT	OEG-4-8-080023-0044-057
NOTE	77p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.95
DESCRIPTORS	*Ability Identification, Aptitude Tests, Cognitive Measurement, Correlation, Evaluation Methods, Grade 10, *Grammar, *Methods Research, Predictive Ability (Testing), Predictive Validity, *Programed Instruction, Program Effectiveness, Research Criteria, Statistical Analysis, *Student Testing, Teaching Techniques, Traditional Grammar, Transformation Generative Grammar
IDENTIFIERS	*Aptitude Treatment Interaction, ATI

ABSTRACT

This research report recounts the procedures, results, and recommendations of a research project in which more than 200 tenth-grade students in Florida were tested (1) to determine whether aptitude treatment interaction (ATI) effects on syntactic maturity and on knowledge of structural relationships occur after several months of instruction; (2) to refine the ability measures used as predictors in this study to increase their differential validity in measuring ATI; (3) to determine whether the findings from the first study could be cross-validated in a second study. The following topics are discussed: (1) previous studies of ATI effects; (2) the complexities involved in any research which attempts to enhance student achievement by using instructional treatments related to ability patterns; (3) the criterion measures used to pretest the students' general, semantic and symbolic abilities and to pre- and post-test their grammar achievement; (4) the effectiveness of the two linear-programed textbooks in transformational and traditional grammar which were used as treatments (5-month period for the first study and 3 months for the second); (5) cross-validation procedures; and (6) the results of the research which, though inconclusive, pointed to modifications in treatments, scoring techniques, and ability tests . (JB)

BR 8-D-023
PA 24

ED0 41896

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Final Report

Project No. 8-D-023
Grant No. OEG-4-8-080023-0044-057

EFFECTIVENESS OF TWO WAYS OF TEACHING GRAMMAR TO STUDENTS OF DIFFERENT ABILITY PATTERNS

F. J. King
Russell P. Kropp
Roy C. O'Donnell

and

William T. Ojala
Michael R. Vitale

Florida State University
Tallahassee, Florida 32306

October, 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TE001 948

PREFACE

The project reported here represents a continuation of efforts to determine whether some students learn more effectively under certain educational treatments while others do better under different treatments. While most people would probably agree that this apparently is true, there have not been sufficient empirical demonstrations of it to provide clear guidelines for its application in classroom situations. The present study was unable to lend strong support to the belief as it applied to the study of grammar by tenth grade students but it demonstrated some of the complexities and to that extent, at least, contributed to the body of knowledge of aptitude treatment interactions.

A number of persons in addition to the investigators contributed to this project. Their assistance is gratefully acknowledged here. Mrs. Nonnie Zeigler, Supervisor of English in the county in which the study was conducted, gave much time and energy to the organization and prosecution of the experiment. The classroom teachers involved, Mrs. Gerri Coggins, Mrs. Annette Flournoy, Mrs. Dorothy Ann Foster, Mrs. Angelia Johnson, and Mrs. Mary Maxwell, were always cooperative and interested in the project.

Dr. William Ojala used a portion of the data collected during this investigation for his dissertation. Several sections of the dissertation were incorporated with little change in this report.

Appreciation is expressed to the staff of the Florida State University Computing Center and to the National Science Foundation for its support of the center.

Finally, appreciation is expressed to the Office of Education, U. S. Department of Health, Education, and Welfare, for its support of the project.

TABLE OF CONTENTS

Page

PREFACE iii

LIST OF TABLES vii

LIST OF FIGURES ix

Chapter

I. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS . . . 1

  Summary 1
  Conclusions 4
  Recommendations 5

II. INTRODUCTION 7

  Problem 7
  Review of Literature 9
  Objectives 20

III. PROCEDURES 23

  Plan of the Research 23
  Tests 26
  Analysis 30

IV. RESULTS 37

  Descriptive Statistics 37
  Results of Tests of ATI Hypotheses 48
  Effectiveness of the Grammar Treatments 62

V. DISCUSSION 65

  Objectives of the Investigation 65
  Limitations of the Study 68

REFERENCES 71

LIST OF TABLES

Number Page

1 The Nine Ability Tests and Their Codes and Reliabilities Used in Grammar Achievement Study 28

2 Means and Standard Deviations of Ability Tests For Both Treatment Groups in Study 1 38

3 Means and Standard Deviations of Ability Tests For Both Treatment Groups in Study 2 38

4 Means and Standard Deviations of Pretest Aluminum Rewrite Scores, Structural Relationships, and Ninth Grade Test Scores for Both Treatment Groups of Study 1. . . . 39

5 Means and Standard Deviations of Pretest Aluminum Rewrite Scores and Ninth Grade Test Scores for Both Treatment Groups of Study 2 39

6 Intercorrelations of SI Ability Measures For Total Group of Study 1 40

7 Intercorrelations of SI Ability Measures For Total Group of Study 2 40

8 Means and Standard Deviations of Criterion Measures for Both Treatment Groups of Study 1 41

9 Means and Standard Deviations of Criterion Measures for Both Treatment Groups of Study 2 41

10 Squared Multiple Correlations for Tests of Linearity of Regression of Posttest on Pretest Aluminum Rewrite Variables 43

11 Squared Multiple Correlations for Tests of Linearity of Regression of Posttest on Pretest Aluminum Rewrite Variables for Control Group 44

12 Means and Standard Deviations of the Aluminum Rewrite Variables for Control Group 44

13 F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models for Words per T-Unit (W/T) in Study 1 49

Number Page

14 F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models for Words per Clause (W/C) in Study 1 50

15 F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models for Clauses per T-Unit (C/T) in Study 1 51

16 Full Model Regression Equations for Treatment by Ability by Pretest Analysis of Study 1 53

17 Numbers and Mean Words Per T-Unit (W/T) Criterion Scores of Subjects Correctly and Incorrectly Classified by the Regression Equation for Correlate Completion 54

18 Numbers and Mean Words Per Clause (W/C) Criterion Scores of Subjects Correctly and Incorrectly Classified by the Regression Equation for Class Name Selection 54

19 Rotated Factor Matrix of Ability Measures for Combined Treatment Groups of Study 1 55

20 F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models in Which Factor Scores Were Used as Independent Variables. . . 56

21 Squared Multiple Correlations for Tests of Linearity of Regression of STEP, Stanford and Structural Relationships on Ninth Grade Variables and Pre Structural Relationships for Study 1 58

22 F-Ratios and Proportions of Variance (R^2's) Accounted for by Full and Reduced Models for STEP and Stanford Tests in Study 1 60

23 Item Selection Data for SI Ability Tests 60

24 Squared Multiple Correlations for the Regression of the Aluminum Rewrite Criteria on Three Versions of Correlate Completion 62

25 Means and Standard Deviations for Three Versions of Correlate Completion 63

26 Mean Words Per T-Unit (W/T) Scores for Subjects of Study 1, of Study 2, and Control Groups in Time Sequence 63

LIST OF FIGURES

Number Page

1 Regression of Mathematics Achievement Scores on Congruent Production of Semantic Transformations (NMT) for Semantic and Symbolic Treatment Group 21

2 Regression of Posttest on Pretest for Words per T-Unit (W/T) 45

3 Regression of Posttest on Pretest for Words per Clause (W/C) 46

4 Regression of Posttest on Pretest for Clauses per T-Unit (C/T) 47

5 Regression of STEP Writing on Math for Study 1 59

I. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The objectives of this investigation were to determine whether aptitude treatment interaction (ATI) effects existed in the study of grammar by tenth grade subjects; whether the results obtained in one study could be reproduced in a second study; whether ATI effects could be enhanced through modification of ability measures; and whether ATI effects identified in the research would have practical implications for educational practice. Two studies were conducted; study 1 sought to identify ATI effects and study 2 attempted to cross validate and intensify the effects found in the first one. In study 1 the subjects were 174 tenth grade English students in six classes in one high school. They were assigned randomly to two treatments which were conducted simultaneously in each classroom. The subjects for study 2 were 115 tenth grade students in five classes from two high schools. Part of them were assigned to treatments by their teachers and part were assigned randomly.

The treatments were two linear programed textbooks in grammar. The first was _English 3200_ and the second was _Modern English Sentence Structure_. A content analysis of the two textbooks indicated that they dealt with the same concepts of the structure of the English language. It also indicated that student success with the first textbook would be mainly related to semantic abilities while student success with the second textbook would involve symbolic abilities in addition to semantic ones. In study 1 the treatment period was five months, and in study 2 it was three months. Few subjects in study 2 managed to complete the treatments in the available time. Seventy-three percent of the subjects in treatment 1 and forty-seven percent in treatment 2 either finished or were in the last quarter of the textbook when the treatment terminated.

For study 1 the criterion measures of major interest, which were also used as pretest variables, were obtained from the _Aluminum Rewrite Test_. These measures were words per T-unit (W/T), words per clause (W/C), and clauses per T-unit (C/T). They were used as indicators of syntactic maturity. Other criterion measures were the STEP _Writing Test_ part 1, _Stanford Achievement Test High School English and Spelling Test_, part B, and the _Test of Recognition of Structural Relationship in English_. The ability measures were nine tests taken from Guilford's Structure of Intellect (SI) battery. They were _Word Classification_, _Verbal Analogies_, _Controlled Association_, _Word Grouping_, _Class Name Selection_, _Number Relations_, _Seeing Trends_, _Memory for Word Classes_, and

Correlate Completion. The first five of these tests were from the semantic content category and the last four from the symbolic content category. The Mathematics and English tests from the Florida State-wide Ninth Grade Testing Program (FSNG-TP) were also used as predictor measures.

In study 2 the criterion measures were the *Aluminum Rewrite Test* variables, which were also used as pretests, and the *STEP Writing Test*. The ability measures were *Word Classification*, *Class Name Selection*, *Seeing Trends*, *Correlate Completion*, and the Mathematics and English FSNG-TP tests. The four SI ability measures were doubled in length, and attempts were made to produce bimodal distributions for them through item selection procedures. The *Aluminum Rewrite Test* was administered twice to thrity randomly chosen tenth grade students who did not participate in either study 1 or 2. The test administrations were two weeks apart. These subjects were used as a non-equivalent control group and their data were used to estimate the reliability of the syntactic maturity variables.

Preliminary analysis of the data of study 1 and of the control group data revealed strong curvilinear relationships between the pretest and posttest *Aluminum Rewrite Test* variables; therefore, quadratic and cubic terms for these variables were included in the analyses which sought to discover ATI effects. Regression models were constructed which contained treatment, ability and pretest variables and their interactions. One model that was used is shown below:

$$\hat{Y} = a + b_1T + b_2Z + b_3X + b_4X^2 + b_5X^3 + (b_6TX + b_7TX^2 + b_8TX^3) + (b_9TZ + b_{10}TZ^2 + b_{11}TZ^3) + (b_{12}TZX + b_{13}TZX^2 + b_{14}TZX^3)$$

T was a dummy variable with values of +.5 and -.5 for treatment 1 (*English 3200*) and treatment 2 (*Modern English Sentence Structure*) respectively. One ability measure was represented by Z, and X was a pretest variable. The full model was fitted, and then reduced models were formed which excluded one set of coefficients (enclosed in parentheses) at a time. The reduction in the squared multiple correlation for each of the reduced models was tested against the full model by an F test.

Significant ATI effects were interpreted by using the regression equation generated by the full model to predict the criterion scores for each treatment for each subject. If the subject's highest predicted criterion score was for the treatment he received, he was called correctly classified. If his highest predicted score was for the treatment he did not receive, he was termed incorrectly classified. If his predicted scores did not differ by more than one-half standard error of estimate, he was considered unclassified. The numbers of subjects who fell into the three classification categories for each treatment and their mean criterion scores were used to determine whether the interactions were disordinal and to indicate the magnitudes of the effects.

In study 2 the cross validation procedures consisted of applying the regression equations obtained in study 1 to the new data, classifying the subjects according to their actual and best predicted treatments, and comparing the mean criterion scores of the resulting groups. Item analyses of the modified ability measures indicated that only Correlate Completion was capable of being modified to give a bimodal distribution. Regression models similar to the one given above were used to compare the magnitudes of ATI effects produced by the bimodal test, the doubled test, and the test as originally used in study 1

The results of the analyses of study 1 data revealed a number of significant ATI effects which involved either the pretest and treatment or an ability, pretest and treatment. These effects typically accounted for two to four percent of the variance of the dependent variable. Comparison of the regression equation for the abilities that produced ATI effects indicated that they were highly similar. In addition, two factor scores derived from the ability measures and interpreted as semantic and symbolic factors also gave highly similar regression equations.

The classification procedure described above was applied to the data of both studies using one regression equation from study 1 for Correlate Completion and one for Class Name Selection where the dependent variables were W/T and W/C respectively. The interactions of study 1 were found to be disordinal and the criterion means were greater for correctly classified subjects than for either incorrectly classified or unclassified subjects. The classification procedure did not yield the expected results when applied to the data of study 2. Failure of the cross validation attempt could be due to the instability of the curved regression planes found in study 1 or to the shorter treatment time in the second study. The latter reason is plausible since both treatment groups in study 1 showed substantial pretest to posttest gains in syntactic maturity but neither group did in study 2. In study 1 the greatest gains were made by subjects in treatment 1. This result was interpreted as being due to more practice in sentence combining by subjects in this treatment.

Analysis of the three versions of Correlate Completion indicated that both the total and bimodal forms produced greater ATI effects than did the original test when C/T was used as the criterion. For all of the Aluminum Rewrite Test criteria the proportions of variance explained by the independent variable were consistently, though not greatly, higher for the revisions of correlate completion than for the original one.

Conclusions

The following conclusions appear to be warranted by the results of the investigation:

(1) Disordinal ATI effects exist in the acquisition of syntactic maturity by tenth grade students but they are relatively weak. They are more complex than originally expected in that they generally involve pretest level of syntactic maturity as well as general ability and treatment. In addition, they involve nonlinear relationships between pre and post measures of syntactic maturity although these relationships are likely to be functions of the instrument used to measure syntactic maturity.

(2) Coefficients of regression models containing significant ATI effects obtained in one study cannot be used to predict the best treatments for subjects of a second study which is of shorter duration than the first. Whether the cross validation failure in this investigation resulted from the fact that many students of the second study failed to complete the textbooks or whether the ATI effects were peculiar to the sample of the first study cannot be determined.

(3) Modification of an ability measure by increasing its spread of scores or by making its distribution bimodal can increase the magnitude of ATI effects in which it is involved.

(4) The present findings concerning aptitude treatment interactions in the acquisition of syntactic maturity and knowledge of structural relationships in English were not sufficiently strong or stable to suggest that tenth grade students could profit from differential placement in one or the other of two grammatical treatments.

Both treatments appeared to accelerate growth in syntactic maturity, but further research should be undertaken to confirm these findings before either treatment is adopted for routine classroom use to achieve this purpose.

Recommendations

(1) Future ATI studies that use syntactic maturity measures as criteria should employ drastically modified treatments that would extend over the entire school year. Each of the present treatments should be changed to include sentence combining exercises similar to those used by Mellon (1967).

(2) Scoring techniques for the Aluminum Rewrite test should be modified to eliminate the nonlinear relationships between the pre and post measures. Concurrent validity studies should be undertaken to determine how well the modified variables predict syntactic maturity as determined from the free writing of students.

(3) Ability tests in future ATI studies of acquisition of syntactic maturity should be measures of general ability. In addition, more specific tests of entering behaviors prerequisite to profiting from the various treatments should be sought. If such tests could be found they might replace either the general ability measure, the pretest of syntactic maturity, or both.

II. INTRODUCTION

Problem

Educational literature contains many references to the desirability of providing for individual differences of students in classroom learning situations. However, the manner in which this individualization of instruction is to be accomplished has usually been unspecified, vague, or based on coarse grouping by achievement or general ability level, and there have not been careful evaluations of the effectiveness of individualized instructional treatments even in those situations where it has been attempted (Carroll, 1963). Thus, little evidence exists that treating a student in one way will cause him to achieve at a higher rate than if he were treated differently. A series of carefully executed studies based on current knowledge of individual differences, learning principles applicable to classroom situations, and statistical and decision theory is needed to determine some ways in which individualization can be accomplished and how much gain in achievement they can produce.

Two of the present investigators have recently completed a series of studies which demonstrated the presence of aptitude treatment interactions in miniature school learning situations (Kropp, Nelson, and King, 1967). These studies, which are reviewed later, indicated that achievement of students can be enhanced by assigning them to instructional materials (treatments) known to be related to their ability patterns. While only a limited number of aptitudes and treatments were examined, the findings implied that it might be possible to tailor for each student a specific curriculum which would take into account his specific ability pattern as well as his learning history.

The present research was designed to refine and extend the findings of the previous studies. Its general purpose was to determine whether aptitude treatment interactions (ATI) persist throughout an extended course of classroom instruction. If interactions occurred only in the early period of instruction, then their practical implications for classroom learning would be limited. If on the other hand they persisted, became more intense, or changed with time, their potential practical value could be demonstrated.

The purposes above were pursued through the study of aptitude treatment interactions in students' acquisition of syntactic maturity and knowledge of structural relationships in English. The treatments were defined by programed textbooks in transformational and traditional grammar. Appropriate ability measures were selected by task analysis procedures.

Reasons for the choice of English grammar as the subject matter area for this investigation were as follows: (1) Recent developments in tranformational grammar suggested that it was sufficiently different from the older grammar to be considered an entirely new treatment. (2) Research studies concerning the effect of traditional grammar on the writing of students have indicated that the traditional techniques are not generally effective. (3) The criterion problem inherent in almost all ATI studies could be minimized. Kropp, Nelson, and King (1967) stated this problem in the following manner:

> One unanticipated difficulty encountered in several of the studies of ATI was the construction of criterion tests that would be appropriate for all treatment groups and adequately assess the common objective of the equivalent sets of materials. An example of this follows and is taken from the mathematics operations studies (in which symbolic and semantic treatments were contrasted). The criterion items for both groups were substantively the same but they were presented in the same form as the instructional materials. If it is claimed that both groups should be able to deal with items in symbolic form, then it is obvious that the semantic group did not reach that objective. In order to reach that objective, the subjects in the semantic group would have to learn the content in verbal form first, then learn the symbol meanings, and then translate their knowledge of mathematical operations into symbolic form during their performance on the criterion. Under these conditions, it is doubtful that the mean criterion performance of the two groups would have been the same. Whether this view of the objective is reasonable depends on determining whether knowledge acquired through the two treatments will transfer equally well to the learning of subsequent material. If equal transfer occurs, then it might be concluded that the equal performance of the two groups on their respective criterion measures implied equal attainment of the objective. If equal transfer does not occur, then special attention should be given to the translation process to determine whether its facilitation could be heightened to preserve the importance of the ATI effect. If it cannot be, then one must conclude that the symbolic treatment is generally the better method of instruction. These points are of great importance in practical application of ATI theory because criterion performance in

> naturally occurring situations usually cannot be altered. Carroll (9) anticipated these difficulties in his suggestion that comparable scores on a criterion test might not be completely indicative of a common goal having been attained. Persons who have been taught by methods that emphasize different abilities and who have achieved comparable criterion scores might differ radically in the way they can use what has been learned in subsequent instructional settings. In addition, treatments that depend on certain abilities might serve to make persons subjected to them more highly differentiated with regard to those abilities and thus predispose them to differential achievement of subsequent instructional goals (Kropp, Nelson, and King, 1967, pp. 204-205).

In the study of grammar, many teachers believe that the most educationally important transfer objective for students is the improvement of their writing skills. An objective measure of syntactic maturity which approximates this criterion exists and is applicable as an outcome measure for any grammatical treatment. (4) Textbooks in programed format exist for both traditional and transformational grammar. The existence of educational materials in different treatment forms is highly important for long term ATI classroom studies. If they were not available, each ATI study would have to become a curriculum writing project before the research could be undertaken.

Review of Literature

An exhaustive review of the literature on ATI was not undertaken here since Cronbach and Snow (1968) have recently completed a USOE report which considered in detail many important ATI studies. Their work integrated the findings, weaknesses, and implications for further research of these studies.

The plan of this section was to review two series of studies that illustrated the nature, intent and outcomes of ATI investigations, and to give an introduction to the Structure of Intellect Model developed by Guilford. Next, the findings and recommendations of Cronbach and Snow that were relevant to the present research were summarized. Finally, studies pertaining to the consequences of grammar instruction were to be reviewed.

Examples of ATI Studies

The two series of studies reviewed below were part of a research program conducted by two of the present investigators (Kropp, Nelson, and King, 1967). The major purposes of the research were (a) to identify form of content variables which exist in instructional materials and which might inhibit or facilitate achievement of these materials by students of different ability patterns; (b) to construct or identify equivalent sets of instructional materials (treatments) that differ in level of one or more form of content variables and (c) to determine empirically the existence of aptitude treatment interactions (ATI) in their achievement by students who vary on relevant ability measures.

The first series of studies dealt with redundancy as a form of content variable. The first two studies of this series used high school students as subjects in relating elements of style to redundancy levels and in identifying cognitive abilities related to redundancy. It was felt that these studies would provide information that would allow the preparation of equivalent sets of material that differed in redundancy levels and the aptitudes which would produce ATI effects. The third and fourth studies used a set of graded-reading material with fifth and sixth grade students and demonstrated that the materials differed in redundancy level. ATI effects were shown for reasoning ability and redundancy levels.

The purposes of the second series of studies were as follows: (a) to determine whether tests of semantic content would be better predictors of achievement of mathematical operations presented in semantic form than they would be of the same materials presented in symbolic form, and whether tests of symbolic content would be better predictors of achievement of symbolic materials than they would be of semantic materials; and (b) to determine the stability and generalizability of ATI effects. Three studies were conducted. The first two used college freshmen as subjects and the third used tenth grade high school students. Materials and tests for all studies were the same. Two sets of materials, one emphasizing semantic abilities and the other emphasizing symbolic abilities, were constructed. Five pairs of ability tests differing only in semantic-symbolic content were used as predictors. In the studies involving college students, the results showed generally stable ATI effects that usually conformed to the theoretical expectations. Many of the results of the study based on high school students were different from the results of the first two studies and from the ATI theory predictions. However, evidence of ATI effects was present.

The Structure of Intellect

The aptitude tests involved in both the studies previously cited and in the present study were taken from the Structure of Intellect Model (SI) (Guilford and Hoepfner, 1963). This three-dimensional model contains a mental process or operation dimension with five categories, a content dimension with four categories, and a product dimension with six categories. The five categories of the operation dimension are cognition (C), memory (M), convergent production (N), divergent production (D), and evaluation (E). The four categories of the content or type of material dimension, the observable stimulus dimension, are figural (F), symbolic (S), semantic (M), and behavioral (B). The six categories of the product dimension, the observable response dimension, are unite (U), classes (C), relations (R), systems (S), transformations (T), and implications (I).

All abilities specified by the model can be identified by a trigram in which the first letter identifies the operation category and the last two letters identify the content and product categories respectively. Thus, CMU, the cognition of semantic units, is the ability that is measured by vocabulary tests. CMR, cognition of semantic relations, is the ability measured by verbal analogies tests.

The SI model is, of course, an imperfect one but it is of value in ATI studies. The 120 abilities that it specifies (excluding the behavioral content category) gives an ample framework for conceptualizing ATI effects.

Review of Cronbach and Snow Report

In a comprehensive review of research in aptitude-treatment interaction, Cronbach and Snow (1968) dealt with the methodology commonly used in such research and found that most of it was quite weak and often inappropriate with interaction effects frequently having been tested by an analysis of variance. As they pointed out, a 2 x 2 analysis of variance, with the subjects divided at the median, does not permit a close look at possible differences in aptitude within either the high or low group. An attempt to divide the group further with high, middle, and low groups, using a 2 x 3 analysis of variance, can "cut the between-groups mean square in half, bringing it below the point of significance." They suggested that regression procedures such as the general linear hypothesis model be employed since tests of ATI effects are essentially tests of homogeneity of regression. Furthermore, since learning is multivariate in nature, analytic techniques should use all information available from sets of both predictor and criterion measures. They also pointed out that interactions between abilities and treatments may be curvilinear although such relationships may be unstable. Finally they suggested that more attention should be paid to descriptive presentations of results.

> Even where the effect in a particular study is not significant, a potential contribution is lost if the effects appearing in the sample are not described. The reality of these weak effects may be more credible if other studies of a similar nature are taken into account. Examination of weak effects also discourages overemphasis on effects within the same study that are not much stronger but that do reach the significance criterion (Cronbach and Snow, 1968, pp. 14-15).

Cronbach and Snow also pointed out that although there have been a number of studies done on specialized abilities, a few have unequivocally demonstrated significant interaction with treatment. Thus, in their comments on the Kropp, Nelson, and King studies, they suggested that

> while a few intercorrelations were negative (between Guilford aptitude tests), there were enough positive intercorrelations to suggest that it would be well to extract one or two main factors, and calculate regression slopes on these more reliable measures. It seems obvious that a fairly general factor would have included most tests of both kinds (symbolic and semantic) and would have entered into a significant interaction, with steeper slopes for the semantic treatment (Cronbach and Snow, 1968, pp. 135-136).

Investigators in this comparatively young field of aptitude-treatment research should, then, begin "by trying to understand just how the general ability complex enters into the learning activities of the pupil."

They also suggested that the treatment or instruction which is under study should be long enough to give the investigator time to determine how the student learns after he has become adjusted to a particular style of instruction. The treatment in this type of research, to be of most value in making educational decisions, should use materials in regular use in educational programs or else some adaptations of them.

Research on the Outcomes of Instruction in Grammar

One report on research on English grammar and composition which had direct bearing on this investigation was prepared at the University of Wisconsin Center for Cognitive Learning (Fredrick, Blount, and Johnson, 1968). In this study, the authors examined the learning of structural grammar by three different modes. Seventy-two eighth graders were randomly assigned to three experimental groups and a control group. The content of all material used with the three experimental groups was the same, emphasizing the concepts of basic sentence, subject group, predicated group, and so on. All three groups used programed lessons. The control group also used a programed text, but the content emphasized poetry reading. The three experimental treatments differed in that the first presented the concepts of grammar entirely in verbal mode, the second in symbolic mode, and the third in figural mode. These different modes of presentation correspond quite closely with Guilford's content dimension with its semantic, symbolic, and figural categories. After five days of practice, all students took a posttest on concepts of English grammar. Two weeks after taking the posttest, the students took a retention test. Student IQ's were obtained from The Lorge-Thorndike Intelligence Tests (Level 4, Form A, Verbal and Nonverbal) which was administered about one and one-half years before this experiment. Those students scoring higher than 116 were put into the high group, those between 105 and 116 were in the medium group, and those below 105 were placed in the low group. Among the results, the authors found that

1. high ability students benefited from all three experimental modes of presenting concepts of English grammar;
2. medium ability students benefited only in the figural and symbolic treatments;
3. low ability students benefited only from symbolic and verbal treatments.

Thus, the authors conclude with these two major generalizations:

1. Learning of grammar concepts can be enhanced through the use of symbols and diagrams,

> provided that the symbols and diagrams are not overly complex for the low ability student.
>
> 2. The presence of the significant interaction between mode of representation and ability suggests that Bruner's concern with matching the mode of representation to the abilities of the learner is entirely warranted. Thus, one must be aware not only of the notation, displays, and models that explicate a subject matter field advantageously, but also of the experience and intelligence of the learner for whom the notation, display, or model must be a vehicle toward understanding rather than a stumbling block (Fredrick, Blount, and Johnson, 1968, pp. 17-18).

Programed learning of structural and transformational grammar and the possible relationships of that learning to writing of eighth grade students was the subject of a study in another report from the same center (Blount, Fredrick, and Johnson, 1968). Over 200 students completed a twenty-two lesson linear programed text on grammar which presented the grammatical concepts of sentence patterns, main structures of a basic sentence, and transforms. Preceding the learning experience, the students submitted 1000 words of free writing; following it they submitted another 1000 word paper. Two types of experimental groups were used: the first followed each lesson in grammar with a worksheet designed to help the students apply what they learned in creating sentences or parts of sentences (Treatment W); the second group did not use this worksheet (Treatment WO). A control group did not study any grammar during the time of the experiment.

The results of this study demonstrated that students in the experimental groups (W and WO) did learn concepts of structural and transformational grammar and that they were able, to some extent, to use the concepts in their writing. However, there were no significant differences between different ability levels on the writing measures. Thus, although brighter students tended to score higher than less bright students on the posttests, they did not demonstrate this difference in their independent writing.

An earlier study comparing the learning of English grammar by means of an automated or programed text and more traditional learning experiences was conducted in the

Denver Public School system (Reed and Hayman, 1962). The major purpose of the study was to find out whether students with average and high achievement learned more than students of low achievement through use of a programed text. About 250 students in five schools participated in the study which covered a period of three months. Students in different "tracks" within the schools were used to determine high, average, and low ability classification. Other measurements, such as IQ, academic rating in English classes, academic rating in other subjects, and scores from three sections of the Iowa Tests of Educational Development, were used as control variables. Two criterion measures were used: the Language Section of the California Achievement Test and a final test taken from the book of tests which accompany English 2600. According to the authors, two pretests were given, one of which was an alternate form of the California Achievement Test, Language Section test; the other was not identified. After the experimental groups had completed the program, the posttests were administered and results analyzed through covariance.

The principal result seemed to be that those students who were in the high ability-high achievement classes did achieve more on both criterion measures, although the experimental groups as a whole did not learn any more than the control group. Furthermore, the control group students classified as low achievers performed better on the criterion tests than experimental group students at the same level. The interaction between programed learning of traditionally presented English grammar and student ability raised the question of whether the same automated instructional materials are suitable for use with students of widely differing academic abilities.

In a somewhat similar study, Bennett (1964) also undertook to discover differences in learning concepts of English grammar through a programed text and the traditional lecture-textbook method of instruction, and the implications of that learning to improved writing. Approximately 120 eleventh-grade students in a Minneapolis high school participated in the two month study. Students were first divided into high, middle and low ability groups on the basis of their performance on the Verbal Reasoning Section of the Differential Aptitude Tests. Those students scoring at the ninetieth percentile and above were designated as the high ability group; those scoring between the sixty-fifth to the ninetieth were designated the middle ability; and those scoring from the sixty-fourth percentile down were designated the low ability group. All students were then assigned randomly to the experimental treatment or to the control group. The experimental treatment consisted of a programed text in traditional grammar, English 3200. The control group also studied traditional grammar, but used either Corbin and Perrin's Guide to Modern English or Warriner and Mersand's English Grammar and Composition. Instruction for the control group followed the same topics given in the programed text. The Cooperative Sequential Tests of Educational Progress--Writing, Form 2A was used as a pretest and also as a posttest. Among his results, Bennett found that

1. programed text and lecture-textbook methods of instruction were equally effective in improving the writing skills of students;
2. programed text seemed to be more effective in teaching grammatical principles and in applying those principles in revising individual, unrelated sentences;
3. there was no interaction between student ability and programed instruction or lecture-textbook instruction, although higher ability students naturally did score higher on both criterion tests.

Bennett also suggested that

> when the goal of instruction is to teach a knowledge of specific grammatical principles and their application to the revision of individual, unrelated sentences, programed instruction in grammar should be used rather than the conventional lecture-textbook method for students of all levels of ability. (Bennett, 1964, p. 67).

This suggestion contradicted the recommendation made by Reed and Hayman that high-ability students do better in programed instruction and low-ability in "traditional learning experiences."

However, two other studies also took differing views on the effectiveness of programed instruction in English grammar with students of differing abilities. The major purpose of the first of these (Kahler, 1966), which was done with tenth grade students, was to

> determine whether programed grammar (English 3200) and/or journal writing would increase student writing ability as measured by an objective, standardized instrument--the Sequential Tests of Educational Progress (Writing Tests) (Kahler, 1966, p. 74-A).

The students were sectioned into high, middle, and low achievement classes based on their past achievement in English classes. One experimental group completed the programed text, English 3200, and a free writing exercise, journal writing. The other experimental group used only the journal writing exercise. A control group given conventional instruction was also employed. The following significant differences were found:

1. low achievement students using both English 3200 and journal writing achieved more than low achievement control group students.
2. middle achievement students in both experimental groups gained more than the control group.
3. high achievement students showed no significant group differences.

Different results were obtained by Hoffman (1968), who hypothesized that tenth and eleventh grade students at all levels of ability and achievement would profit from the study of materials in a programed instruction format supplemented with present methods in English grammar. Six classes of tenth grade students and five classes of eleventh grade students were randomly assigned to receive the experimental program or to serve as the control group. Two criterion measures were used: the Cooperative English Test (English usage section) and a mastery test used with the programed texts, English 2600 (used with the tenth grade students) and English 3200 (eleventh grade students). Both criterion tests also served as pretests and as retention tests. Further, the students were divided into "quartiles" according to their ability and achievement. Hoffman found that the programed instruction was more effective with students in the highest quartile and less effective with students in the lowest quartile.

One other study on the use of programed texts in learning English grammar is of interest here. Using twelfth grade students, Munday (1966) had his experimental group complete English 3200 while the control group was taught conventionally, through use of a non-programed text, drill, lecture, and mimeographed sheets. One of his conclusions was that students who completed the programed textbook learned as well as those who had conventional instruction and that they did so in a shorter period of time. Also, they retained about as much information from their learning as did students in a conventional situation.

Two studies having to do with improvement of student writing through instruction in transformational grammar had implications for the present project. The first of these was done by Bateman and Zidonis (1966). Four specific questions guided this study:

1. Can high school pupils learn to apply the transformational rules of a generative grammar in their writing?
2. Can their repertoire of grammatical structures be increased by a study of generative grammar?
3. To what extent will the proportion of well-formed sentences increase in pupil writing over the two-year period?
4. What kinds of transformational errors will occur in pupil writing, and to what extent will such errors increase or diminish over the two-year period? (Bateman and Zidonis, 1966, p. 3).

Data from forty-one students ultimately were analyzed after the experimental group had received instruction in generative grammar over a two-year period. This instruction came in addition to the regular curriculum completed by both experimental and control groups. Samples of student writing were taken during the first three months of the experiment and during the last three months; these constituted before and after samples. The writing was analyzed primarily for (1) types of transformations used in the student sentences; (2) structural complexity score; (3) proportion of well-formed sentences; and (4) error change score. Although the results were frequently ambiguous, the authors claimed to have shown that a relation exists between knowledge of generative grammar and the ability to produce well-formed sentences of great structural complexity, and further, that knowledge of generative grammar enabled the experimental subjects to increase sentence complexity without sacrificing grammaticality. There did not seem to be any correlation between student IQ and the amount of increase in the proportion of well-formed sentences as measured on posttest examination.

The other investigation was done by Mellon (1967), who studied the influence of grammar-related sentence-combining practice on student syntactic fluency. Two-hundred and forty seven seventh grade students were assigned to three treatment groups: transformational sentence-combining, conventional parsing, and no grammar. The students came from urban, suburban, and private schools, and represented five levels of ability. Pre and post writing was done in response to nine parallel topics assigned during the fall and again during the spring. This writing was then parceled off into T-units and analyzed according to twelve factors of syntactic fluency, mainly nominal and relative embeddings, frequency and depth of embedding, and clustered modification. Mellon's major result was that the experimental group, who had the sentence-combining practice, made gains in syntactic fluency significantly greater than the control group. The possible difference in gain in syntactic fluency due to ability level was not, however, so apparently clear. Even though there were three significant F-ratios among the twelve factors of syntactic fluency, Mellon was properly cautious in attributing the interaction to the difference in abilities:

> There is some question whether the significant interactions should be attributed more to the regression tendency of the controls, or more to the offsetting tendency of the experimental treatment to exert its uniformly positive effect to a degree that is proportionate to initial developmental standing, and thus differentially (Mellon, 1967, p. 98).

He did, however, make this conclusion on the effect of ability interacting with the sentence-combining problems:

> While the occurrence of growth was uniform within the experimental group regardless of whether subject ranked in the upper or lower half of the group on the scale of pre-practice development, it can be argued, although somewhat ambiguously, that the magnitude of this growth was significantly greater for the initially high-half subjects than for those in the low half, as compared with growth observed in the high and low halves of the control group (Mellon, 1967, p. 107).

Pertinent literature reviewed here demonstrates that interactions between certain abilities and certain instructional treatments can be obtained under experimental conditions. It also indicates that although different treatments exist for the teaching of English grammar, no investigations of their interactions with abilities of students have been carried out. However, studies cited do give some evidence that such interactions can be produced.

Objectives

The specific objectives of the proposed study were as follows: (1) to determine whether ATI effects on syntactic maturity and on knowledge of structural relationships in English occur after several months of instruction; (2) to modify and refine the ability measures used as predictors in order to increase their differential validity in measuring ATI; (3) to replicate the study, if time permitted, in order to determine whether ATI effects are consistent and whether the revised ability measures are better indicators of ATI than the unrevised ones; and (4) if ATI effects were discovered, to conduct utility studies to determine whether the increased cost of using two kinds of instructional materials outweighs the increments in learning that they produce.

An example of ATI effects on student achievement, taken from the second series of ATI studies cited in the review of literature is shown below. Heterogeneous groups of students were taught certain mathematical operations involved in vector multiplication and in the computation of the derivative of an algebraic expression by either highly symbolic or highly verbal methods. An achievement test whose items differed only in form (symbolic or verbal) was administered to each group. Figure 1 shows the graphs of the regression equations of achievement on one of Guilford's tests for the ability factor, convergent production of semantic transformations (NMT), for each group.

The crossover point of the regression lines occurs at score 13 of the test of NMT. Thus, if maximum achievement of the concepts taught in this study were desired, students below score 13 on NMT should be assigned to the symbolic method and students above score 13 should be given the verbal (semantic) treatment. To be more specific, the equations predict that a student with a score of 5 will achieve a score of 25 if taught by the symbolic method but he will achieve a score of only 18 if taught verbally. Conversely a student with a score of 16 will achieve 28 on the verbal materials but only 24 with the symbolic treatment.

If similar results were to be found in achievement of important educational objectives then decisions of the value of the achievement increments would need to be made.

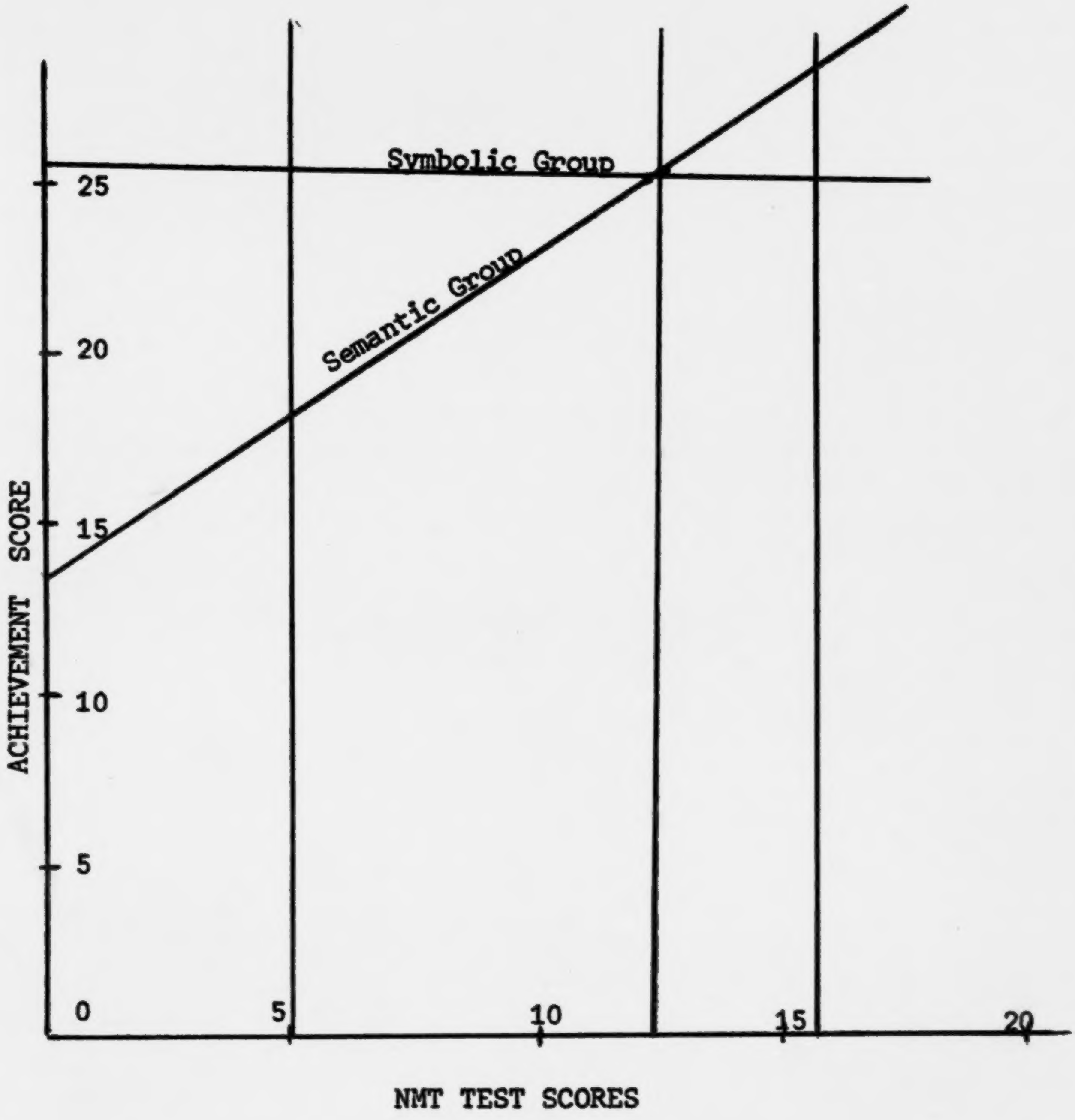


Figure 1. Regression of Mathematics Achievement Scores on Congruent Production of Semantic Transformations (NMT) for Semantic and Symbolic Treatment Groups

III. PROCEDURES

Plan of the Research

The original plan for the research project called for the two treatments to be administered in the spring semester to students in the tenth grade, for the data to be analyzed in the summer, and for a partial replication study to take place in the fall. This schedule would have allowed a thorough analysis of the data from the first study to be made during the summer. Ample time for test selection and revision for the second study would have been available.

Because of delays of the starting date of the project, the first study was begun in the fall. In addition, the first study consumed more time than was anticipated. Consequently the time for analysis of data and decisions about the instruments to be used in the second study was shorter than was desirable.

Subjects

The subjects for study 1 were taken from a high school in a small North Florida community. The community is located in a predominately agricultural area with shade grown leaf tobacco and some beef cattle the principal products. The community has approximately 9,000 persons living in it. There are two high schools, one of which is primarily Negro; the other school, in which this study took place, is integrated. With the exception of a few wealthy families, the community could be considered lower-middle to middle class economically.

All of the students in the tenth grade English classes participated in the study; there were 174 students in six classes. However, due to absences on days tests were scheduled, separate analyses in this study contain differing numbers of students. Four of the classes were taught by one teacher; the other teacher involved had two tenth grade English classes. The students were assigned at random to the two treatments which were conducted simultaneously within each class. Eighty-seven students were assigned to each treatment.

The subjects for study 2 came from two high schools in two communities located within the same county as the school in which study 1 was done. In one school one teacher taught two classes of twenty-five and twenty-seven students, and one teacher taught one class of seventeen students. In the other school one teacher taught two classes of seventeen and twenty-nine students. In the first school the investigators found that the teachers, in preparing to participate in the study, had carefully divided their students on the basis of prior class performance into two groups for the two treatments. Because of the time pressures involved and because of the work which the teachers had performed, it was decided to use these groups rather than to insist on random assignment. In the second school the students were randomly assigned to treatments.

An additional factor complicated the assignment of students to treatment. Fewer copies of the programed text for treatment 1 were recovered from study 1 than were texts for treatment 2. Therefore, the randomly assigned students were unequally divided. The total number of students, 115, was broken into two treatment groups containing 51 and 64 students for treatment 1 and 2 respectively. The treatments were conducted within each classroom simultaneously. Again some students were absent on some days on which tests were administered so that analyses of the data contain different sample sizes.

Instructional Treatments

In both studies the treatments were two linear programed textbooks in grammar. The first was English 3200 (Blumental, 1962); the other was Modern English Sentence Structure (Rogovin, 1964). A content analysis by two faculty members of the Department of English Education at Florida State University indicated that the greater part of both books deal with the same concepts of the English language and that they would require approximately the same amount of time to complete. English 3200 is a traditionally-oriented textbook, presenting its descriptions and explanations almost wholly in verbal form. It does not utilize the Reed-Kellogg diagraming of sentences common to most traditional grammar texts. Because of its high verbal content and its relative lack of use of symbols and diagrams, it was considered to emphasize abilities that lie in the semantic category of Guilford's Structure of Intellect model.

Rogovin's Modern English Sentence Structure is based on the transformational-generative description of the structure of English grammar. Thus, it uses a great number of symbols such as NP, VP, S, ------>, and tree diagrams to explain concepts of grammar and the relationships found between elements of the sentence. Because of its heavy use of symbols and tree diagrams, this textbook was considered to emphasize abilities that lie at least partially in the symbolic category of the Structure of Intellect model.

This analysis of both textbooks agreed rather well with that done by Fredrick, Blount, and Johnson (1968). Although the Modern English text used in this study does contain tree diagrams, which Fredrick, Blount, and Johnson considered to be figural presentations, there are many more rewrite rules and other explanations which employ symbolic material than there are tree diagrams. Therefore, even though this text does contain figural presentations of grammatical relationships, it was considered for this study to be primarily symbolic in content.

Even though the texts were programed, the teachers were an important part of the treatment. All of the teachers involved in both studies had taken at least one course in transformational grammar and were favorably disposed toward it. They did have some reservations about the effectiveness of programed instruction.

At the beginning of the school year before the first study began, three group meetings were held with all teachers. The first was a general orientation session concerning the project, and the others, conducted by a professor of English Education, dealt with the content and use of the programed texts.

The general instructions given to the teachers were that the textbooks should be used in a way consistent with their perceptions of good teaching. They were encouraged to work with individuals or small groups of students when they felt it was needed and to vary classroom activities when they felt it was desirable. They were encouraged to use the unit tests provided by the publisher for English 3200 for checking the progress of students and for their own grading purposes. Similar tests for groups of units for Modern English Sentence Structure were constructed by the investigators to be used in the same way. The first five of these tests were made available to the teachers in multiple copies. The remaining seven, because of time pressures, were available only in one copy to each teacher. This resulted in less than optimal usage of these tests.

Two important differences between the two studies occurred with regard to instructional procedures. Since the textbooks were to be used twice, the students in the first study were required to use a mimeographed answer sheet for recording their responses. The students in the second study recorded responses in their books. The separate answer sheets proved to be a source of confusion and annoyance to both teachers and pupils and probably increased the time spent in the treatment. The second difference was that initially the students of the first study were not allowed to take the textbooks out of the classroom. It was felt that better control over their work would be achieved in this way and that textbook loss would be minimized. This restriction was later relaxed in order to speed student progress through the programs. The students in the second study were allowed to take their textbooks home from the onset.

The first study began on September 17, 1968, with the administration of ability tests and pretests. The program was completed on February 19, 1969; when the last posttest was given. Both groups began their treatment by studying in the programed texts for all five periods of the week. However, the classroom teachers felt that the schedule allowed them no time to give to other aspects of English, particularly the reading of literature. Thus, approximately two weeks after the program began, the schedule was changed to have the students work in their programed texts three days of the week and to study literature the other two days. Both groups went through the same content in their literature study. All study by both groups in the programed grammar texts was done during the class periods.

Because the syllabus for the tenth grade English class required the study of Julius Caesar, work in class by both groups was interrupted for the period of time required to go through this play. Upon returning to their programed texts, the students resumed their three-period-a-week schedule; in addition, they were allowed to take the texts out of the class to work at home and in study halls.

The second study began on February 21, 1969, and terminated on May 20, 1969; the end of the school term. Thus the first study consumed approximately five months of time while the second was conducted in only three. Not all students in the latter study completed the textbooks. Records of the units or lessons which each student had completed were made at approximately monthly intervals.

Tests

The aptitude tests used in the first study were chosen from among those available at the time to represent possible differences of ability in the symbolic and semantic content categories. These choices reflected the differences between the semantic and symbolic contents of the two programed texts. Since the grammar of a language concerns itself with classes of elements, such as noun phrases, verb phrases, adjectives, pronouns, etc., tests were chosen which would reflect that concern. Furthermore, not only does grammar concern itself with classes of elements of a language, but it also describes and explains relationships of elements within the language. Therefore, from the product dimension of Guilford's three-dimensional model, those tests were chosen which lie within the categories of classes and relations.

Since it was believed that the student responds to new material first by trying to comprehend it in the form in which it is presented to him, four out of the nine ability tests chosen lie within the operations category of cognition. Two tests lie within the convergent production category. This category encompasses the type of reasoning which is most frequently used, wherein the individual attempts to take given information and from it generate more information in a unique (to the individual) but conventionally accepted form. In the case of programed texts in grammar, the student is given grammatical information from which he is asked to generate information not necessarily having variety or quantity. One test apiece from the operation categories of memory, divergent production, and evaluation were chosen to fill out the dimension of operations. The nine ability tests, arranged so that a symbolic content test would alternate with a semantic content test, were assembled in two booklets. These tests, with their Guilford codes and reliabilities, are listed in Table 1.

Two tests, also used as criterion measures, were administered before the students began the program. The first of these was the Aluminum Rewrite Test which was developed by Dr. Roy C. O'Donnell for a study done by Hunt (1968). It is a passage consisting of 32 sentences of connected discourse on the subject of aluminum. Each sentence in the passage is extremely short, averaging about 4 1/3 words; each sentence is a single independent clause. Students were given the passage and asked to rewrite it "in a better way." The rewritten versions were then analyzed for number of words, number of clauses, and number of T-units. Having all students write in response to the same passage and in the same mode of exposition allowed analyses of the rewritten passage to be made for comparative purposes across groups.

Syntactic maturity can be measured by analyzing written passages into the ratios of words per T-unit (W/T), words per clause (W/C), and clauses per T-unit (C/T). In a series of studies by Hunt (1964, 1965, 1966) and in a similar study by O'Donnell, Griffin, and Norris (1967), the T-unit was developed

Table 1

The Nine Ability Tests and Their Codes and Reliabilities Used in Grammar Achievement Study

	Code	Test	r
1.	CMC	Word Classification	$.76^a$
2.	CMR	Verbal Analogies	$.80^a$
3.	DMR	Controlled Association	$.82^b$
4.	NMC	Word Grouping	$.75^c$
5.	EMC	Class Name Selection	$.63^d$
6.	CSC	Number Relations	$.81^a$
7.	CSR	Seeing Trends II $.80^e$	$.56^a$
8.	MSC	Memory for Word Classes	
9.	NSR	Correlate Completion II	$.76^a$

a Davis (1967)

b Guilford, Kettner, and Christensen (1954)

c Guilford, Merrifield, Christensen, and Frick (1960)

d. Nihira, Guilford, Hoepfner, and Merrifield (1964)

e. Hoepfner, Guilford, and Merrifield (1964) and expanded into a highly effective measurement of syntactic maturity. Briefly, the term "T-unit" means a "minimal terminable unit." It is the shortest unit which can be separated from other units without producing a sentence fragment. Hunt's definitions of sentence, clause, and T-unit were used in this analysis: Sentence--"the words written between a capital letter and a period or other terminal punctuation;" clause--"a structure containing a subject (or coordinated subjects) and a finite verb phrase (or coordinated verbs or phrases);" T-unit--"one main clause plus the subordinate clauses attached to or embedded within it."

The second test which was used as pre-posttest was a Test of Recognition of Structural Relationships in English. The test was developed by Roy C. O'Donnell (1963) and used by him in a study measuring the relationship between knowledge of structural relationships and written composition. The test has 50 items of the three-option multiple-choice type. It was designed to measure ability in recognizing structural relationships in the English language. The items were written in such a way as to avoid use of formal grammatical terminology; recognition of predication, complementation, coordination, modification, and cross reference were measured. Administering the test to over 100 high school seniors, O'Donnell obtained a split-half reliability coefficient (Spearman-Brown formula) of .88; inter-item consistency coefficient (Kuder-Richardson formula) of .86.

Sections of two standardized tests were used as additional criterion measures. Part 1 of the STEP Writing Test consists of four separate short compositions. The student is asked to read each one and then to answer several questions on the passage. The questions are of the four-option multiple-choice type; there are 30 items in Part 1. Although some of the questions deal with punctuation and spelling, most of them are questions of sentence effectiveness, clarity, vividness, and paragraph structure. This test is concerned with the writing of whole compositions, and so applications of grammatical principles are all placed in the context of a piece of writing several paragraphs in length. This test also avoids the use of grammatical terminology which would undoubtedly tend to favor one group of students over the other.

The second test used was Part B of the Stanford Achievement Test, Form W, High School English and Spelling Tests, which consists of ten items of the four-option multiple-choice type. Each item asks the student to respond by choosing the one sentence which best expresses an idea.

The results from a battery of tests administered in September, 1967, to the students used in this study when they were beginning the ninth grade, were also available for some analyses during this study. This battery is the Florida State-Wide Ninth-Grade Testing Program. There are five tests in the battery: 1) Aptitude, 2) Social Studies, 3) English, 4) Mathematics, and 5) Science. The scores used in this study were the aptitude, English, and mathematics tests.

The pretests used in the second study were chosen on the basis of a preliminary correlation analysis of the study 1 data. Tests that appeared to have higher correlations with either STEP, Stanford or Sentence Relations in one group than the other were selected. It would have been desirable to base the selection on more thorough regression analyses and on use of the Aluminum Rewrite Criterion data, but time did not permit this approach. The tests that were selected were as follows:

Aluminum Rewrite
Class Name Selection
Correlate Completion
Seeing Trends
Word Classification

The four tests from the SI Model were doubled in length by the inclusion of new items written by members of the project staff. The new items were similar to those of the form given in study 1. They were made into separately timed tests and were given after the part administered in study 1.

The criterion tests used in study 2 were the Aluminum Rewrite and STEP Writing Test, Part 1. Structural Relationships was not included since no mean pre-post changes occurred in study 1. The Stanford subtest was eliminated because of low ceiling effects.

Analysis

Previous studies that have had as their primary concern the investigation of ATI effects have frequently compared regression slopes of dependent on independent variables for two treatments. If more than one independent variable was being studied, the analysis considered the pair of slopes for each variable separately. In studies that investigated ATI only incidentally, the independent variable often was used to produce levels in a treatment-by-levels design. Cronbach and

Snow pointed out that ATI studies which include more than one independent variable could probably profit by considering them all within one analysis. When one independent variable is a pretest, it should be included in the analysis as an aptitude rather than being used to form a gain score. They also suggested that non-linear models may yield more information than linear models, but cautioned against overfitting and against uncritical acceptance of weights for non-linear terms in a regression equation.

In the studies reviewed by Cronbach and Snow and in their own work they failed to find impressive evidence of differential aptitudes interacting with treatments. Most ATI effects seemed to them to be the result of general ability and they suggested that many of the studies they reviewed would have been improved by using the first factor from a set of independent variables as the aptitude measure to be analyzed.

Because the analytic procedures for studying ATI have not been well defined, the methods used for analyzing the data of this investigation included many of the techniques recommended by Cronbach and Snow. However, some of the simpler methods of previous studies were also utilized.

Analyses of Data in Study 1

In analyzing the results of the Aluminum Rewrite posttest the following procedures were used:

1. A preliminary analysis was made to determine whether the relationship between pretest and posttest scores was non-linear. It was believed that non-linearity might occur for all three scores (words per T-unit [W/T], words per clause [W/C], and clauses per T-unit [C/T]) because of floor and ceiling effects. The full model had the form

$$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3 \qquad (1)$$

where $\hat{Y}$ is the estimated posttest score and X is the pretest. It was evaluated by testing the difference between R^2's of $\hat{Y} = a + b_1X$ vs. $\hat{Y} = a + b_1X + b_2X^2$ and the difference between $\hat{Y} = a + b_1X + b_2X^2$ vs. the full model. The two comparisons tested the quadratic and cubic effects respectively. The comparisons were made for each treatment group separately and for the groups combined. The results did show evidence of non-linearity so the linear and quadratic terms were carried in subsequent analyses.

2. The next series of analyses involved an aptitude variable (Z), the pretest scores (X, X^2, X^3), and the treatments (T). One analysis was conducted for each of the nine aptitudes. In the following model, T is a dummy variable with values of +.5 and -.5 respectively for Treatment 1 (English 3200) and Treatment 2 (Modern English):

$$\hat{Y} = a + b_1T + b_2Z + b_3X + (b_4X^2 + b_5X^3) + (b_6TX + b_7TX^2 + b_8TX^3) + (b_9TZ + b_{10}TZ^2 + b_{11}TZ^3) + (b_{12}TZX + b_{13}TZX^2 + b_{14}TZX^3) \qquad (2)$$

The full model was fitted, and then reduced models were employed which excluded one set of coefficients (enclosed in parentheses) at a time. The reduction in R^2 for each of the reduced models was tested against the full model. If, for example, the reduced model excluding ($b_6TX + b_7TX^2 + b_8TX^3$) produced a significant F-ratio when tested by the formula

$$F = \frac{R^2_{\text{Full}} - R^2_{\text{Reduced}}}{1 - R^2_{\text{Full}}} \cdot \frac{N - M_1 - 1}{M_1 - M_2}$$

(where M_1 is the number of variables in the full model, M_2 is the number of variables in the reduced model, and N is the total sample size), the interaction produced by the effect of the treatment by linear, quadratic and cubic terms of the pretest was declared to be significant.

3. Interpretations of significant interaction effects found in the analyses above were made by using the regression equation generated by model 2 to predict the criterion score for each subject for each treatment. If the difference between the predicted scores was greater than one-half of the standard error of estimate of the equation, the subject was classified as "belonging" to the treatment that produced the highest predicted score. If the difference between predicted scores was equal to or less than one-half standard error of estimate, the subject remained unclassified. The actual treatment that each subject received was then identified, and he was placed in one of the six categories shown in the diagram that follows:

	Predicted Treatment		
Actual Treatment	T_1 (English 3200)	Unclassified	T_2 (Modern English)
T_1			
T_2			

The numbers of subjects falling into the six categories allowed the determination of the type of interaction involved. If both predicted treatment categories contained subjects, then a disordinal interaction existed, i.e. the regression planes crossed within the ranges of the independent variables. If one predicted treatment contained no subjects, then the interaction was ordinal, i.e. the slopes of the planes were different but did not intersect within the range of scores of the independent variables. The means of the actual criterion scores of the subjects in the six categories, coupled with the number of subjects in each category, gave an indication of the magnitude of the effect. It was expected that mean criterion scores of subjects correctly classified would be greater than those of unclassified subjects and incorrectly classified subjects would have the lowest means of all.

4. The aptitude variables were factor analyzed by the principal components method and rotated by varimax procedures. Factor scores were computed by regression for the first two factors and their scores were studied for ATI effects. This seemed to be a desirable procedure in view of the findings of Cronbach and Snow and in view of the similar results given by many of the aptitude measures in the analyses mentioned above. The use of factor scores rather than individual aptitude measures is parsimonious, and the possibility of their producing greater ATI effects than the aptitude variables is present since the factor scores should be more reliable than any of the single measures alone. The individual factor scores were analyzed according to model (2), except that the $TZ^2 + TZ^3$ terms were dropped, where Z was first factor one and then factor two. In addition, the two factor scores were analyzed together with the pretest in the following model:

$$\begin{aligned}\hat{Y} = {} & a + b_1T + b_2Z_1 + b_3Z_2 + b_4X + (b_5X^2 + b_6X^3) + \\ & (b_7TX + b_8TX^2 + b_9TX^3) + (b_{10}TZ_1 + b_{11}TZ_2 + b_{12}TZ_1Z_2) + \\ & (b_{13}TZ_1X + b_{14}TZ_1X^2 + b_{15}TZ_1X^3) + \\ & (b_{16}TZ_2X + b_{17}TZ_2X^2 + b_{18}TZ_2X^3) \qquad (3)\end{aligned}$$

The reduced models formed by eliminating sets of coefficients in parentheses were tested against the full model.

The analyses of the remaining dependent variables (STEP Writing Test, Part 1, Stanford Achievement Test, Part B, and the Structural Relationships Posttest) were accomplished in a manner similar to those of the Aluminum Rewrite.

1. A preliminary analysis for each dependent variable was made to determine whether its relationship to each of the ninth grade scores (aptitude [verbal and quantitative], English, and math total) and the pretest Structural Relationships was non-linear. The model used was that given in (1). A number of non-linear components were found to be significant, so they were included in the remainder of the analyses.

2. Treatment by independent variable interactions were analyzed for the STEP and Stanford by the following model:

$$\begin{aligned}\hat{Y} = {} & a + b_1T + b_2X + (b_3X^2 + b_4X^3) + \\ & (b_5TX + b_6TX^2 + b_7TX^3) \qquad (4)\end{aligned}$$

where X was math total for two of the analyses and English for the other two. In addition, both math total and English were used in the model given below to determine whether they would have either separate or joint ATI effects on the STEP, Stanford, and Structural Relationships tests. For these analyses, Z stands for math total and X for English.

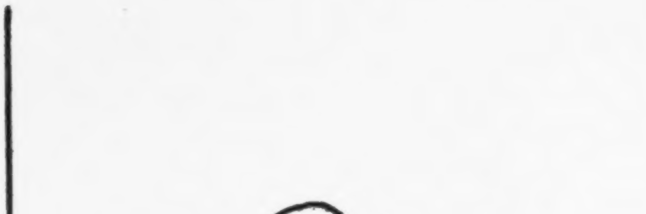
$$\begin{aligned}\hat{Y} = {} & a + b_1T + b_2X + b_3Z + b_4X^2 + b_5X^3 + b_6Z^2 + b_7Z^3 \\ & + (b_8TX + b_9TX^2 + b_{10}TX^3) \\ & + (b_{11}TZ + b_{12}TZ^2 + b_{13}TZ^3) \\ & + (b_{14}TXZ + b_{15}TX^2Z^2 + b_{16}TX^3Z^3) \qquad (5)\end{aligned}$$

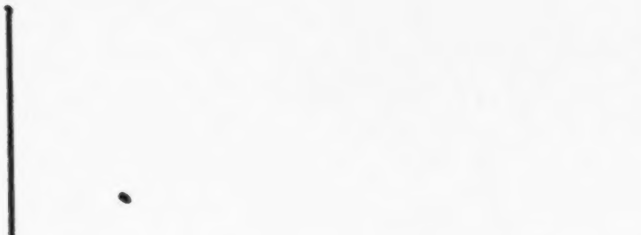
3. Since pretest data on *Structural Relationships* were present, it and math total were used to test for ATI effects on the posttest *Structural Relationships* data. Model (2) was used for this analysis. A similar analysis for English was not made since in none of the other analyses did it show evidence of producing ATI effects.

4. Factor scores for the aptitude variables were included in separate analyses with math total or English to determine whether they would have separate or joint ATI effects on the STEP test. Stanford and *Structural Relationships* were not considered since no significant ATI effects had occurred for them in previous analyses. The model used for these analyses was (5), where X stands for the respective factor score and Z stands for math total or English.

Analyses of Data in Study 2

The analyses of the data gathered in study 2 fall into two main sets. The first set consisted of attempts to replicate the results of study 1. The classification procedure previously described was used on study 2 data with the regression equations derived from the appropriate models and data of study 1. The second set of analyses consisted of attempts to refine the SI ability tests to increase any possible ATI effects. The two parts of each SI test were intercorrelated so that reliability estimates could be obtained. All of the item difficulties and intercorrelations for each test were examined. It was felt that if bimodal distributions for the tests could be produced, ATI effects would be enhanced by them if the crossover points of the regression lines or planes for the treatment groups occurred at points of the distribution where low frequencies occurred. The diagrams below illustrate this reasoning.





The diagrams show hypothetical frequency distributions for an ability variable in normal and bimodal forms and the regression lines for the criterion variable for each treatment. The left one indicates that most subjects' scores occur around the crossover point. This is the place where classification decisions are most uncertain. If the modified test maintained the same regressions for the two treatments as shown in the right diagram, then many more classification decisions could be made and the ATI effect would be magnified.

IV. Results

Almost all of the analyses of study 1 were completed before those of study 2 were undertaken. In this chapter, however, the results of both studies are presented together in order to make their similarities and differences more apparent. The first section of this chapter gives a summary of the descriptive statistics obtained in both studies; the next section presents the results of the tests of the ATI hypotheses. The final section presents findings concerning the general effectiveness of the grammatical treatments.

Descriptive Statistics

The means and standard deviations of the SI ability tests for both treatment groups for study 1 are given in Table 2. English 3200 is represented by T_1, and T_2 represents Modern English Sentence Structure. The means and standard deviations of the SI ability measures used in study 2 are shown in Table 3. The parts of the measures that were the same as those of study 1 are labeled "1"; the parts that were constructed during the project are labeled "2".

The means and standard deviations of the remainder of the pretest measures are presented in Tables 4 and 5 for studies 1 and 2 respectively. The intercorrelations of the SI ability measures for the total groups are presented in Tables 6 and 7 for the two studies. The data presented in Tables 2 through 7 indicate that the subjects in both studies were quite similar and that the treatment groups within studies were initially comparable. Tables 8 and 9 show the means and standard deviations of the criterion tests for each treatment group for the two studies.

Again the groups appeared to be roughly comparable although greater pre-post differences seemed to occur in study 1 for the Aluminum Rewrite variables than in study 2. In addition, pre-post differences appeared to be greater for T_1 in study 1 and greater for T_2 in study 2. It should be noted here that all of the subjects in study 2 did not complete the textbooks. For English 3200 (T_1),73% of the subjects either had finished the book or were in the last quarter of it, 18% were in the third quarter, and 9% were in the second quarter when the treatment terminated. For Modern English Sentence Structure (T_2), 47% either had finished or were in the last quarter of it, 42% were in the third quarter, and 11% were in the second quarter when the treatment terminated.

Table 2
Means and Standard Deviations of Ability Tests For Both Treatment Groups in Study 1

		T_1			T_2		
Test		M	SD	N	M	SD	N
Controlled Associations	(DMR)	12.84	7.54	77	11.36	6.15	69
Number Relations	(CSC)	11.96	4.34	77	13.10	4.43	69
Word Classifications	(CMC)	9.96	2.87	77	9.88	2.67	69
Correlate Completion II	(NSR)	7.13	5.30	77	5.80	5.19	69
Verbal Analogies	(CMR)	9.50	2.53	78	9.35	3.05	72
Memory for Word Classes	(MSC)	24.53	7.61	78	26.84	8.04	72
Word Grouping	(NMC)	19.83	6.87	78	18.68	7.91	72
Class Name Selection	(EMC)	9.83	2.43	78	10.60	2.09	72
Seeing Trends	(CSR)	2.64	1.99	78	2.49	2.08	72

Table 3
Means and Standard Deviations of Ability Tests For Both Treatment Groups in Study 2

	T_1			T_2		
Test	M	SD	N	M	SD	N
Word Classification 1	9.08	2.75	50	10.07	2.84	57
Word Classification 2	9.96	3.17	50	10.26	2.78	57
Correlate Completion 1	9.41	5.91	49	8.75	6.28	59
Correlate Completion 2	6.37	4.94	49	6.92	5.63	59
Class Name Selection 1	10.58	3.32	50	11.23	3.86	57
Class Name Selection 2	10.06	2.50	50	10.42	2.47	57
Seeing Trends 1	2.48	1.89	46	2.49	1.92	57
Seeing Trends 2	1.85	1.84	46	2.00	1.77	57

Table 4

Means and Standard Deviations of Pretest Aluminum Rewrite Scores, Structural Relationships, and Ninth Grade Test Scores for Both Treatment Groups of Study 1

Test	T_1			T_2		
Aluminum Rewrite	M	SD	N	M	SD	N
Words per T-Unit (W/T)	9.43	2.23	73	9.25	2.16	68
Words per Clause (W/C)	6.92	.99	73	7.05	1.15	68
Clauses per T-Unit (C/T)	1.35	.26	73	1.32	.24	68
Structural Relationships	21.77	6.49	77	20.67	6.06	69
Aptitude	60.71	17.72	79	58.36	18.21	76
English	42.47	12.61	79	39.87	13.97	76
Mathematics	46.82	13.80	79	46.59	15.13	75

Table 5

Means and Standard Deviations of Pretest Aluminum Rewrite Scores and Ninth Grade Test Scores For Both Treatment Groups of Study 2

Tests	T_1			T_2		
Aluminum Rewrite	M	SD	N	M	SD	N
Words per T-Unit (W/T)	9.26	2.93	46	9.49	2.70	55
Words per Clause (W/C)	6.79	1.26	46	6.91	1.13	55
Clauses per T-Unit (C/T)	1.35	.32	46	1.37	.31	55
Aptitude	64.22	11.70	31	61.47	16.48	43
English	43.58	10.40	31	40.98	10.53	43
Mathematics	48.39	12.35	31	46.93	15.61	43

Table 6

Intercorrelations of SI Ability Measures For Total Group of Study 1

Test	1	2	3	4	5	6	7	8	9
1. Controlled Associations		.48	.49	.55	.47	.33	.37	.44	.54
2. Number Relations			.48	.56	.36	.22	.32	.23	.44
3. Word Classification				.48	.43	.19	.38	.38	.40
4. Correlate Completion					.48	.27	.41	.34	.66
5. Verbal Analogies						.27	.45	.39	.35
6. Memory for Word Classes							.11	.21	.34
7. Word Grouping								.43	.33
8. Class Name Selection									.39
9. Seeing Trends									

Table 7

Intercorrelations of SI Ability Measures For Total Group of Study 2

Test	1	2	3	4	5	6	7	8
1. Word Classification 1		.47	.42	.42	.22	.40	.40	.42
2. Word Classification 2			.64	.61	.36	.57	.46	.49
3. Correlate Completion 1				.80	.58	.51	.57	.61
4. Correlate Completion 2					.58	.49	.55	.65
5. Class Name Selection 1						.53	.47	.48
6. Class Name Selection 2							.57	.50
7. Seeing Trends 1								.66
8. Seeing Trends 2								

Table 8

Means and Standard Deviations of Criterion Measures For Both Treatment Groups of Study 1

Test		T_1			T_2	
Aluminum Rewrite	M	SD	N	M	SD	N
Words per T-Unit (W/T)	10.26	2.23	73	9.76	2.24	68
Words per Clause (W/C)	7.20	1.29	73	7.07	1.15	68
Clauses per T-Unit (C/T)	1.44	.23	73	1.40	.27	68
STEP Writing Part I	14.66	5.80	74	14.42	5.35	62
Stanford Part B	7.47	3.38	73	7.39	2.21	64
Structural Relationships	21.90	6.37	73	20.57	6.91	65

Table 9

Means and Standard Deviations of Criterion Measures For Both Treatment Groups of Study 2

Test		T_1			T_2	
Aluminum Rewrite	M	SD	N	M	SD	N
Words per T-Unit (W/T)	9.37	2.13	46	9.75	2.30	55
Words per Clause (W/C)	6.92	1.16	46	7.04	1.27	55
Clauses per T-Unit (C/T)	1.35	.24	46	1.38	.21	55
STEP Writing Part I	15.00	5.93	49	14.76	5.11	59

Preliminary analyses of the *Aluminum Rewrite* variables were made to determine whether the relationship between pretest and posttest scores was non-linear, to determine test-retest reliability of the scores and to investigate the sensitization effects of pretesting. In order to help make these analyses, the *Aluminum Rewrite* test was administered twice to five tenth grade English classes who did not participate in the experiment. The test administrations were two weeks apart. Thirty subjects who had taken both tests were randomly selected and used as a non-equivalent control group for the treatment groups in studies 1 and 2, for investigating reliability, and for studying sensitization effects. One of the thirty subjects produced unscorable papers so that the group was reduced to twenty nine.

Table 10 shows the results of the tests of linearity of regression of posttest on pretest for the three *Aluminum Rewrite* variables. Similar information for the control group is given in Table 11. The means and standard deviations on the *Aluminum Rewrite* scores for the control group are shown on Table 12. Tables 10 and 11 show strong non-linear effects for all three variables although the results are not highly consistent. The strong second and third degree effects for word per clause (W/C) in the control group indicate that the non-linearity is a function of the test itself and not of the treatments involved in the two studies. The same conclusion might also be true for words per T-unit (W/T). It is likely that the quadratic term would have been significant if more subjects had been included in the control group. The lack of relationship of pre and posttest for the clauses per T-unit (C/T) scores in the control group is somewhat surprising since its standard deviations are similar to those of the treatment groups. However, the pre-post relationship fluctuates rather markedly within the treatment groups themselves so that no inference about control group-treatment group differences can be made.

The forms of the curvilinear relationships for the *Aluminum Rewrite* scores for all groups are shown in Figures 2, 3, and 4. The shapes of the relationship for all groups are similar for W/T but quite different for W/C and C/T. Even for W/T the relationships of the curves of the treatment groups within studies are different. Thus, in study 1 treatment 1 produced the highest predicted scores for subjects who scored high on the pretest; in study 2 the situation was reversed.

All curves for all groups, with one exception (W/C for study 1), give estimated posttest scores less than actual pretest scores for initially high pretest performance. Predicted increments occur only at the lower to middle end of the pretest

Table 10

Squared Multiple Correlations for Tests of Linearity of Regression of Posttest on Pretest Aluminum Rewrite Variables

Independent Variables	Treatments - Study 1					
		T_1			T_2	
	W/T	W/C	C/T	W/T	W/C	C/T
X	.412	.326	.064	.299	.114	.150
$X + X^2$	.450*	.351	.191*	.355*	.370*	.163
$X + X^2 + X^3$	.450	.408*	.191	.453*	.398*	.230*
	Treatments - Study 2					
		T_1			T_2	
	W/T	W/C	C/T	W/T	W/C	C/T
X	.513	.256	.448	.385	.439	.351
$X + X^2$	.618*	.378*	.450	.562*	.448	.430*
$X + X^2 + X^3$	.619	.380	.464	.566	.449	.436

*Increase in R^2 over previous R^2 significant beyond .05.

Table 11

Squared Multiple Correlations for Tests of Linearity of Regression of Posttest on Pretest Aluminum Rewrite Variables for Control Group

Independent Variables	W/T	W/C	C/T
X	.425	.278	.038
$X + X^2$	.486	.511*	.040
$X + X^2 + X^3$	.490	.632*	.050

*Increase in R^2 over previous R^2 significant beyond .05

Table 12

Means and Standard Deviations of the Aluminum Rewrite Variables For Control Group

	Pretest		Posttest	
Variable	M	S.D.	M	S.D.
W/T	9.54	2.81	9.72	2.45
W/C	6.96	1.30	7.05	1.29
C/T	1.36	.26	1.38	.29

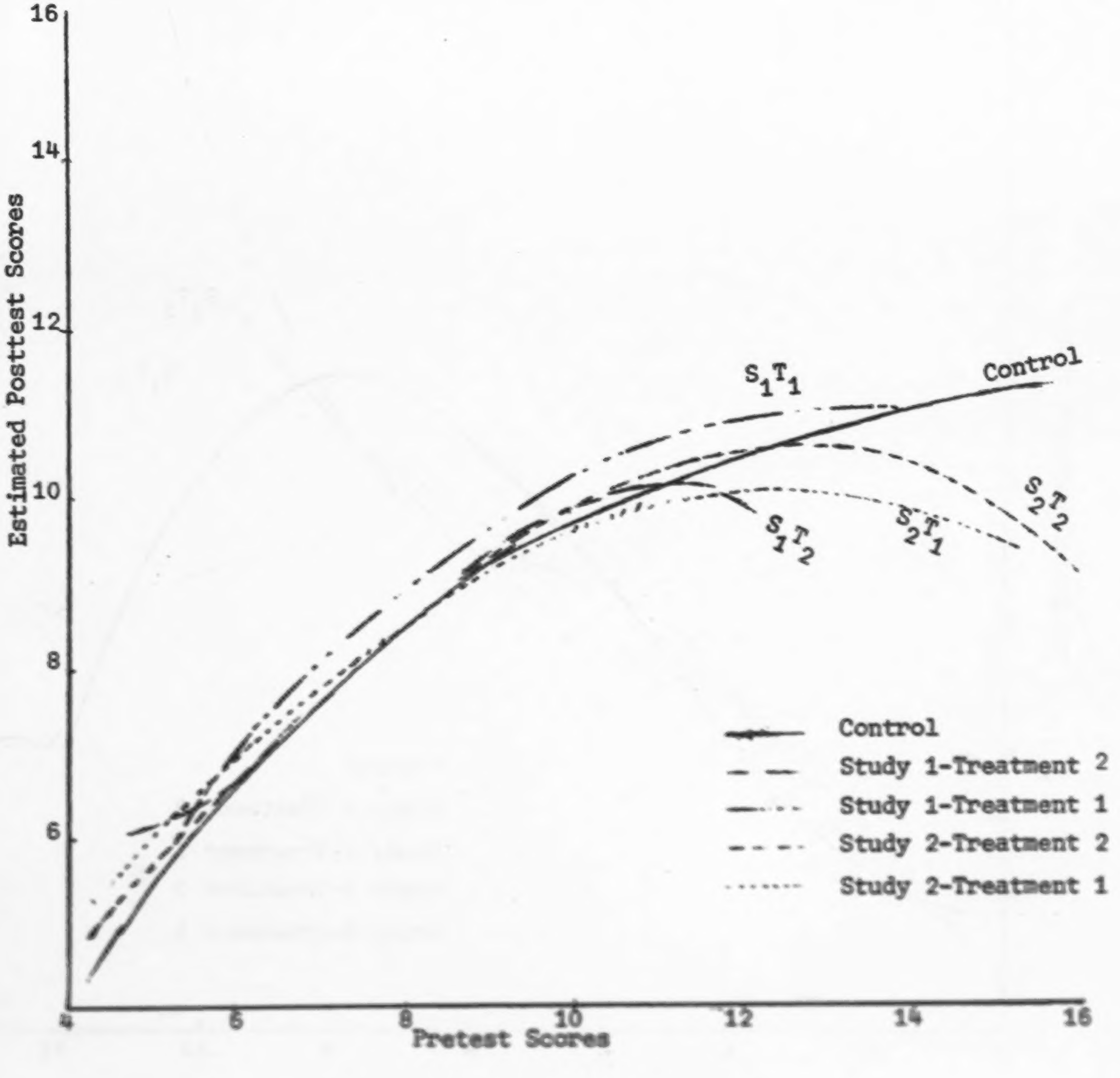


Figure 2. Regression of Posttest on Pretest for Words per T-Unit (W/T)

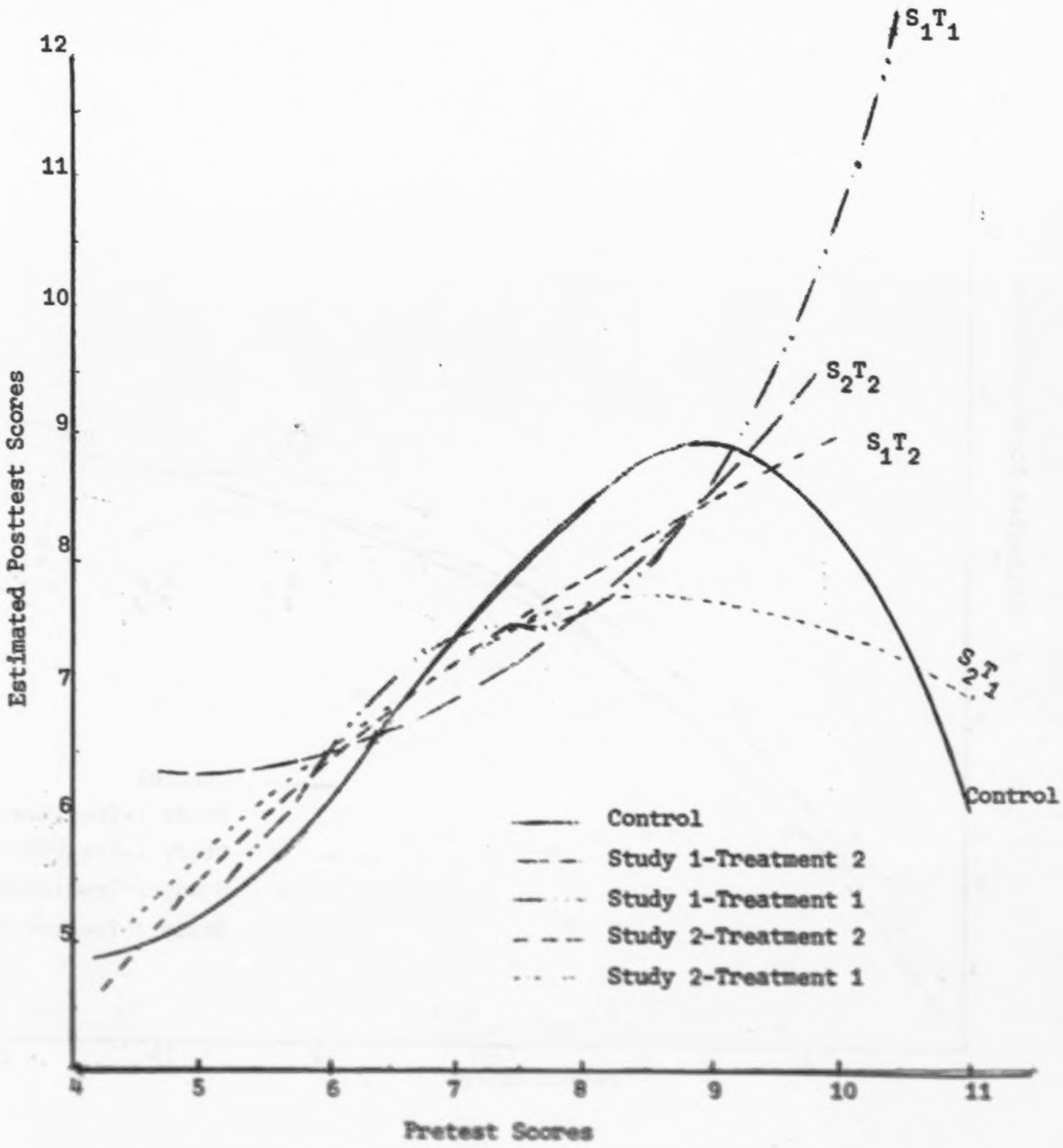


Figure 3. Regression of Posttest on Pretest for Words per Clause (W/C)

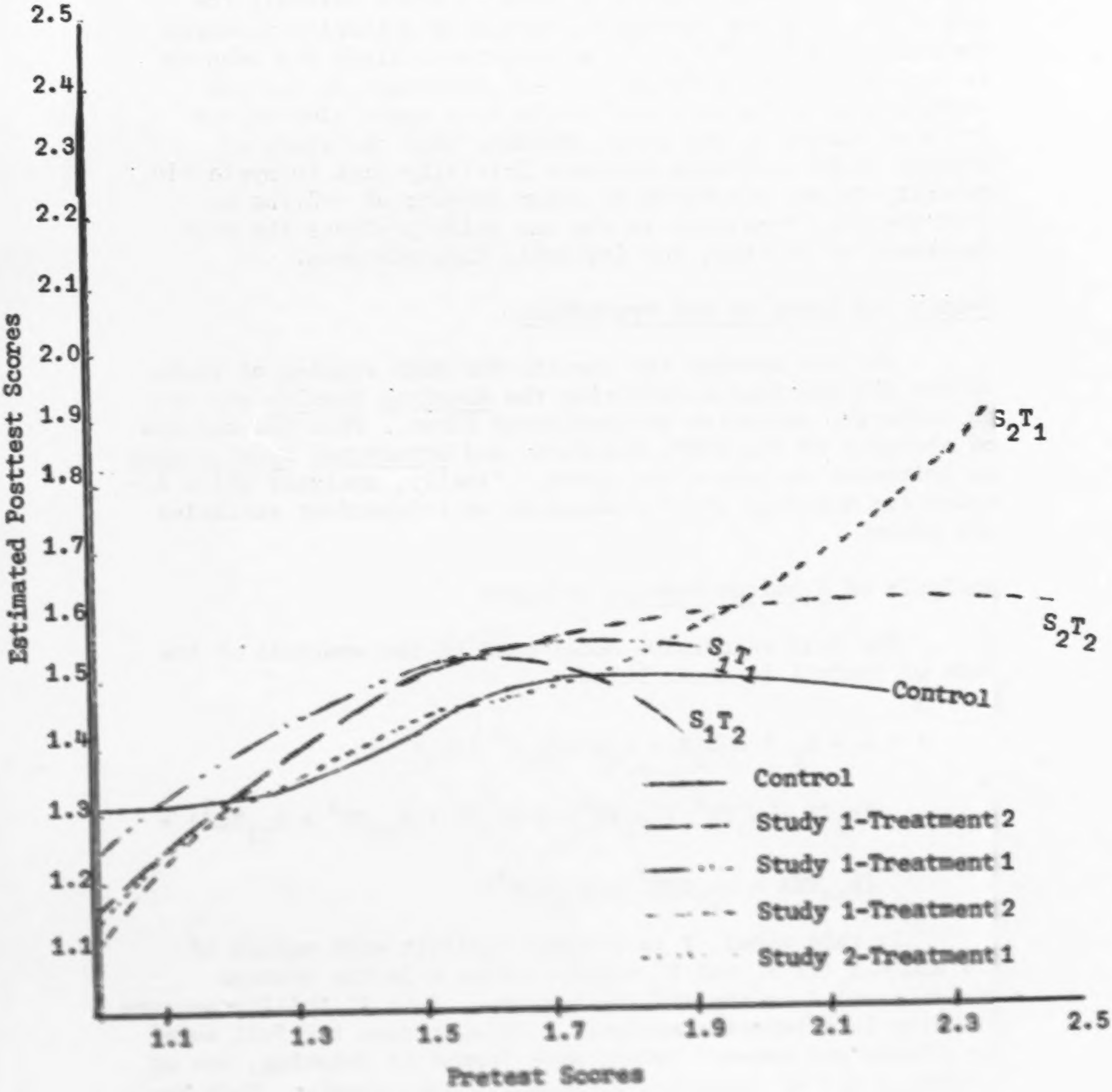


Figure 4. Regression of Posttest on Pretest for Clauses per T-Unit (C/T)

distribution. This finding appears to limit severely the use of the *Aluminum Rewrite* variables as criterion measures for assessing the value of the treatments since its meaning is unclear. Most probably the best treatment is the one which produces the most resistance to a regression effect for high subjects, but it is possible that the study of grammar might encourage students initially high in syntactic maturity to pay attention to other aspects of writing so that the best treatment is the one which produces the most decrement at posttest for initially high subjects.

Results of Tests of ATI Hypotheses

In this section the results for both studies of tests of the ATI hypotheses involving the *Aluminum Rewrite* scores as criterion variables are presented first. Then the results of analysis of the STEP, Stanford, and *Structural Relationship* as criterion variables are given. Finally, analyses which involve the modified ability measures as independent variables are shown.

Analysis of *Aluminum Rewrite* Criteria

The full regression model used in the analysis of the data of study 1 is shown below:

$$\hat{Y} = a + b_1 T + b_2 Z + b_3 X + b_4 X^2 + b_5 X^3 +$$

$$(b_6 TX + b_7 TX^2 + b_8 TX^3) + (b_9 TZ + b_{10} TZ^2 + b_{11} TZ^3) +$$

$$(b_{12} TZX + b_{13} TZX^2 + b_{14} TZX^3)$$

In this model, T is a dummy variable with values of +.5 and -.5 for T_1 and T_2 respectively; X is the pretest variable and Z is the ability measure. Each SI ability measure was used in a separate analysis. In each case the full model was fitted and reduced models were formed by deleting, one at a time, a set of variables enclosed in parenthesis. Each reduced model was tested according to the formula given in the analysis section to determine whether the deleted set of variables contributed significantly to the prediction of the dependent measure. The results of these analyses for each of the criterion variables W/T, W/C, and C/T are shown in Tables 13, 14, and 15 respectively.

Inspection of these tables reveals that none of the ability measures interacted significantly with the treatment alone for any of the criterion measures. Significant

Table 13

F-Ratios and Proportion of Variance (R^2's)
Accounted for by Full and Reduced Models
For Words Per T-Unit (W/T) in Study 1

Test		Full Model	$TX+TX^2 + TX^3$	$TZ+TZ^2 + TZ^3$	$TZX+TZX^2 + TZX^3$
1. Class Name Selection (EMC)	R^2	.5788	.5532	.5721	.5537
	F		2.5530	.6680	2.5030
2. Word Classification (CMC)	R^2	.5494	.5258	.5420	.5281
	F		2.1997	.6900	1.9850
3. Correlate Completion (NSR)	R^2	.5693	.5440	.5571	.5409
	F		2.4671	1.1896	2.7694*
4. Seeing Trends (CSR)	R^2	.5589	.5452	.5563	.5477
	F		1.3044	.2480	1.0660
5. Controlled Associations (DMR)	R^2	.5479	.5341	.5453	.5355
	F		1.2820	.2415	1.1520
6. Word Grouping (NMC)	R^2	.5683	.5294	.5658	.5268
	F		3.7845*	.2430	4.0375*
7. Memory for Word Classes (MSC)	R^2	.5406	.5312	.5301	.5338
	F		.7960	.9600	.6216
8. Verbal Analogies (CMR)	R^2	.5599	.5283	.5480	.5293
	F		3.0156*	1.1356	2.9202*
9. Number Relations (CSC)	R^2	.5807	.5515	.5616	.5527
	F		2.9248*	1.9131	2.8046*

*Significant at .05, N = 141

Table 14

F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models For Words per Clause (W/C) in Study 1

Test		Full Model	$TX+TX^2 + TX^3$	$TZ+TZ^2 + TZ^3$	$TZX+TZX^2 + TZX^3$
1. Class Name Selection (EMC)	R^2	.5124	.4732	.4947	.4695
	F		3.3765*	1.5246	3.6952*
2. Word Classification (CMC)	R^2	.4958	.4766	.4795	.4775
	F		1.5993	1.3577	1.5243
3. Correlate Completion (NSR)	R^2	.4943	.4937	.4683	.4840
	F		.0498	2.1593	.8554
4. Seeing Trends (CSR)	R^2	.4904	.4894	.4647	.4713
	F		.0824	2.1181	1.5741
5. Controlled Associations (DMR)	R^2	.4864	.4808	.4803	.4781
	F		.4579	.4988	.5787
6. Word Grouping (NMC)	R^2	.4607	.4540	.4558	.4538
	F		.5217	.3816	.5373
7. Memory for Word Classes (MSC)	R^2	.4645	.4588	.4600	.4628
	F		.4470	.3529	.1333
8. Verbal Analogies (CMR)	R^2	.4875	.4854	.4779	.4868
	F		.1720	.7867	.0573
9. Number Relations (CSC)	R^2	.4985	.4703	.4860	.4653
	F		2.3617	1.0468	2.7888*

*Significant at .05, N = 141

Table 15

F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models For Clauses Per T-Unit (C/T) in Study 1

Test		Full Model	$TX+TX^2+TX^3$	$TZ+TZ^2+TZ^3$	$TZX+TZX^2+TZX^3$
1. Class Name Selection (EMC)	R^2	.2980	.2922	.2885	.2911
	F		.3470	.5683	.4128
2. Word Classification (CMC)	R^2	.3138	.2967	.3034	.2965
	F		1.0466	.6365	1.0588
3. Correlate Completion (NSR)	R^2	.3416	.3316	.3147	.3298
	F		.6379	1.7159	.7527
4. Seeing Trends (CSR)	R^2	.3052	.2924	.2944	.2887
	F		.7737	.6528	.9974
5. Controlled Associations (DMR)	R^2	.3041	2904	.2970	.2875
	F		.8268	.4285	1.0018
6. Word Grouping (NMC)	R^2	.3008	.2803	.2979	.2841
	F		1.2314	.1741	1.0031
7. Memory for Word Classes (MSC)	R^2	.3091	.3007	.2811	.3039
	F		.5106	1.7021	.3161
8. Verbal Analogies (CMR)	R^2	.3368	.3064	.3184	.3086
	F		1.9[illegible]52	1.1652	1.7858
9. Number Relations (CSC)	R^2	.3362	.3064	.3131	.3073
	F		1.8855	1.4615	1.8285

Significant at .05, N = 141

interactions of *Correlate Completion*, *Word Grouping*, *Verbal Analogies*, and *Number Relations* with the pretest and treatments were found for W/T. Significant pretest by treatment interactions were found in the presence of *Word Grouping*, *Verbal Analogies*, and *Number Relations* for W/T. For the W/C criterion only *Class Name Selection* and *Number Relations* produced significant ability by pretest by treatment interactions. A significant pretest by treatment interaction occurred only in the presence of *Class Name Selection*. No significant interactions of any kind were found for the C/T criterion. The regression equations for ability variables that showed significant interactions for the criterion variables W/T and W/C are shown in Table 16.

The patterns of magnitudes and signs of the regression coefficients for all five equations are highly similar. Therefore, only two of them were selected for further analysis and interpretation on the assumption that the other three would yield highly similar results. The two equations that were selected were the ones that employed *Correlate Completion* and *Word Classification* as ability measures for the W/T and W/C criterion measures respectively.

For each regression equation the classification procedure described in the analysis section was employed in order to interpret the interactions. Tables 17 (*Correlate Completion*) and 18 (*Word Classification*) show the results of the classification procedure for study 1 and also the cross validation results for study 2 which were obtained by applying regression coefficients of study 1 to the study 2 data.

The data of Tables 17 and 18 show that both interactions are disordinal; that is, the regression planes intersect within the ranges of the independent variables. However, the relatively small number of subjects with predicted T_2 classifications probably indicates that the crossover lines lie near the ends of the distributions. In Table 17 the means for study 1 have the expected relative magnitudes. The correctly classified subjects have a higher mean than the unclassified subjects for whom no treatment can be predicted to be best. The unclassified subjects have a higher mean than the subjects who were incorrectly classified. Study 1 data in Table 18 show that the mean for correctly classified subjects is greater than the others but that the mean is greater for incorrectly classified subjects than for unclassified ones.

The cross validation procedure using the regression equations of study 1 and the data of study 2 did not produce the predicted results for either criterion measure. The

Table 16

Full Model Regression Equations for Treatment by Ability by Pretest Analysis of Study 1

Source of Variation	Criterion Measure				
	W/T Ability Measure				W/C Ability Measure
	Correlate Completion	Word Grouping	Verbal Analogies	Number Relations	Class Name Selection
Intercept (a)	-10.766	-18.151	- 7.601	-12.345	- 12.208
Treatment (T)	-27.320	- 7.688	-80.607	-52.761	-247.167
Ability (Z)	.024	.010	- .001	.055	.074
Pretest (X)	4.926	7.170	3.736	5.359	6.399
X^2	- .391	- .610	- .246	- .444	- .755
X^3	.011	- .018	.006	.013	.032
TX	11.046	5.401	32.581	18.667	114.389
TX^2	- 1.376	- .939	- 4.046	- 2.314	- 17.393
TX^3	.045	.046	.158	.088	.882
TZ	4.300	1.110	6.117	5.262	25.353
TZ^2	- .051	- .005	.111	- .126	.054
TZ^3	.002	.000	- .004	.003	- .001
TZX	- 1.558	- .469	- 2.724	- 1.411	- 11.393
TZX^2	.185	.066	.336	.172	1.714
TZX^3	- .007	- .002	- .013	- .006	- .085

Table 17

Numbers and Mean Words Per T-Unit (W/T) Criterion Scores of Subjects Correctly and Incorrectly Classified by the Regression Equation for Correlate Completion

Actual Treatment		Treatment Classification		
		T_1	Uncertain	T_2
Study 1				
T_1	N	34	29	10
	M	11.23	10.15	7.28
T_2	N	28	37	3
	M	10.21	9.31	10.62
T_1+T_2	N	62	66	13
	M	10.77	9.68	8.05
Study 2				
T_1	N	29	8	8
	M	9.91	7.87	8.71
T_2	N	31	18	5
	M	10.74	8.54	7.31
T_1+T_2	N	60	26	13
	M	10.34	8.33	8.17

Table 18

Numbers and Mean Words Per Clause (W/C) Criterion Scores of Subjects Correctly and Incorrectly Classified by the Regression Equation for Class Name Selection

Actual Treatment		Treatment Classification		
		T_1	Uncertain	T_2
Study 1				
T_1	N	23	43	7
	M	7.80	6.95	6.80
T_2	N	26	34	8
	M	7.32	6.68	7.95
T_1+T_2	N	49	77	15
	M	7.55	6.83	7.41
Study 2				
T_1	N	14	24	7
	M	7.29	6.97	6.00
T_2	N	17	27	5
	M	7.76	6.82	6.32
T_1+T_2	N	31	51	12
	M	7.55	6.89	6.13

incorrectly classified subjects tend to have means which are equal to or higher than those of the subjects in the other groups. The failure of the cross validation procedure can probably be attributed to the instability of the curved regression planes and to the fact that the crossover lines occur at the extremes of the ranges of independent variables. In addition, the lack of random assignment of subjects to treatments and the shorter treatment duration in study 2 which did not allow all subjects to finish their textbooks may have contributed to the cross validation failure.

The SI ability variables for the subjects of the combined treatment groups in study 1 were factor analyzed by the principle components method. Two factors were rotated by the varimax procedure. The rotated factor matrix is presented in Table 19. Factor one appears to be defined primarily by the semantic tests and factor two by the symbolic ones. However, only a few of the tests are relatively pure measures of either factor.

Table 19

Rotated Factor Matrix of Ability Measures For Combined Treatment Groups of Study 1

Ability Test	Factor 1	Factor 2
1. Controlled Associations (DMR)	.501	.605
2. Number Relations (CSC)	.380	.591
3. Word Classification (CMC)	.559	.418
4. Correlate Completion II (NSR)	.486	.651
5. Verbal Analogies (CMR)	.645	.323
6. Memory for Word Classes (MSC)	-.116	.749
7. Word Grouping (NMC)	.804	.042
8. Class Name Selection (EMC)	.702	.160
9. Seeing Trends II (CSR)	.361	.701

Factor scores computed by regression were obtained for each subject of study 1. They were used in the same full and reduced regression models as the individual ability measures, except that the treatment by ability measure squared and cubed terms were omitted.

The results of the tests of the sets of interaction variables are shown in Table 20. The regression equations were highly similar to those of the individual SI ability measures so that no further analyses for purpose of interpretation were made. Since only four of the ability measures were used in study 2, factor scores could not be computed and no cross validation attempts could be made.

One further analysis was made which used both factor scores and the pretest, their squared and cubed terms, and their interactions as independent variables. The proportions of variance accounted for by the full model was only slightly larger than the proportions of variance reported in Table 20. No interaction effects that were markedly different from those of the prior analyses were found.

Table 20

F-Ratios and Proportion of Variance (R^2's) Accounted for by Full and Reduced Models in Which Factor Scores Were Used as Independent Variables

Criterion	Factor		Full Model	$TX + TX^2 + TX^3$	TZ	$TZX + TZX^2 + TZX^3$
W/C	1	R^2	.4791	.4722	.4728	.4694
		F		.5600	1.5200	.7800
	2	R^2	.4741	.4658	.4740	.4738
		F		.6800	.0200	.0200
W/T	1	R^2	.5849	.5495	.5846	.5497
		F		3.5800*	.0900	3.5600*
	2	R^2	.5648	.5295	.5524	.5320
		F		3.4100*	3.5900*	3.1700*

*Significant at .05; N = 139

Although it was not the purpose of this study to investigate treatment main effects, it is of interest that in none of the analyses were significant treatment effects obtained. The pretest was always significantly related to the criterion, and the ability effects generally were significant.

Analysis of STEP, Stanford and *Structural Relationships*

A preliminary analysis was made for each treatment group of study 1 to determine whether nonlinear relationships existed between the variables of the ninth grade test battery and the STEP, Stanford, and *Structural Relationships* (SR). The results presented in Table 21 give evidence of nonlinear relationships which differ from one treatment to the other. It appeared desirable, therefore, to include the quadratic and cubic terms of the ninth grade variables and SR (pre) in the investigations of the ATI hypothesis.

The following model was used to determine ATI effects on the STEP and Stanford tests:

$$\hat{Y} = a + b_1T + b_2X + b_3X^2 = b_4X^3 = (b_5TX + b_6TX^2 + b_7TX^3)$$

In one pair of analyses X represented the Math variable, and in another pair it represented English. The ninth grade aptitude score was not included in these or other analyses because the preliminary analysis of nonlinear effects indicated that it would give results which would be very similar to those of the English score. The results of the analyses involving Math and English, which are given in Table 22, show that only the treatment by Math interaction for the STEP variables was significant. No significant interactions for the Stanford test were found. The classification procedure previously described was applied to the STEP data of both study 1 and study 2. The results of these analyses indicated that the interaction was an ordinal one since only two subjects were classified as being able to profit most from treatment two. The regression lines for each treatment for the Math by treatment interaction for study 1 are shown in Figure 5. The regression equation from which they were computed is shown below:

$$T = 17.39 + 14.13T - .79M + .02M^2 - .00M^3 - 1.11TM + .02TM^2 - .00TM^3$$

The regression lines for the two treatments indicate that subjects who are in the lower part of the distribution of Math scores do better on the STEP writing test when they study *English 3200* (T_1) than when they study *Modern English Sentence Structure* (T_2). Students in the upper part of the distribution of Math

Table 21

Squared Multiple Correlations for Tests of Linearity of Regression of STEP, Stanford and Structural Relationships on Ninth Grade Variables and Pre Structural Relationships for Study 1

Ninth Grade Variables	T_1			T_2		
	STEP	Stan	SR	STEP	Stan	SR
Aptitude (X)	.350	.230	.199	.578	.452	.236
$X + X^2$	.391*	.261	.238	.611*	.465	.337*
$X + X^2 + X^3$	.399	.329*	.254	.616	.468	.361
English (X)	.472	.320	.277	.608	.510	.328
$X + X^2$	.533*	.337	.330*	.637*	.512	.559*
$X + X^2 + X^3$	.563*	.471*	.342	.644	.527	.559
Math (X)	.379	.262	.262	.303	.391	.061
$X + X^2$	.391	.263	.326*	.462*	.401	.341*
$X + X^2 + X^3$	.402	.331	.327	.587*	.604*	.391*
SR (Pre) (X)	.343	.106	.534	.283	.207	.371
$X + X^2$	.343	.115	.547	.308	.210	.491*
$X + X^2 + X^3$	.343	.117	.549	.313	.224	.499

*Increase in R^2 over previous R^2 significant beyond .05 N for T_1 = 67; N for T_2 = 57.

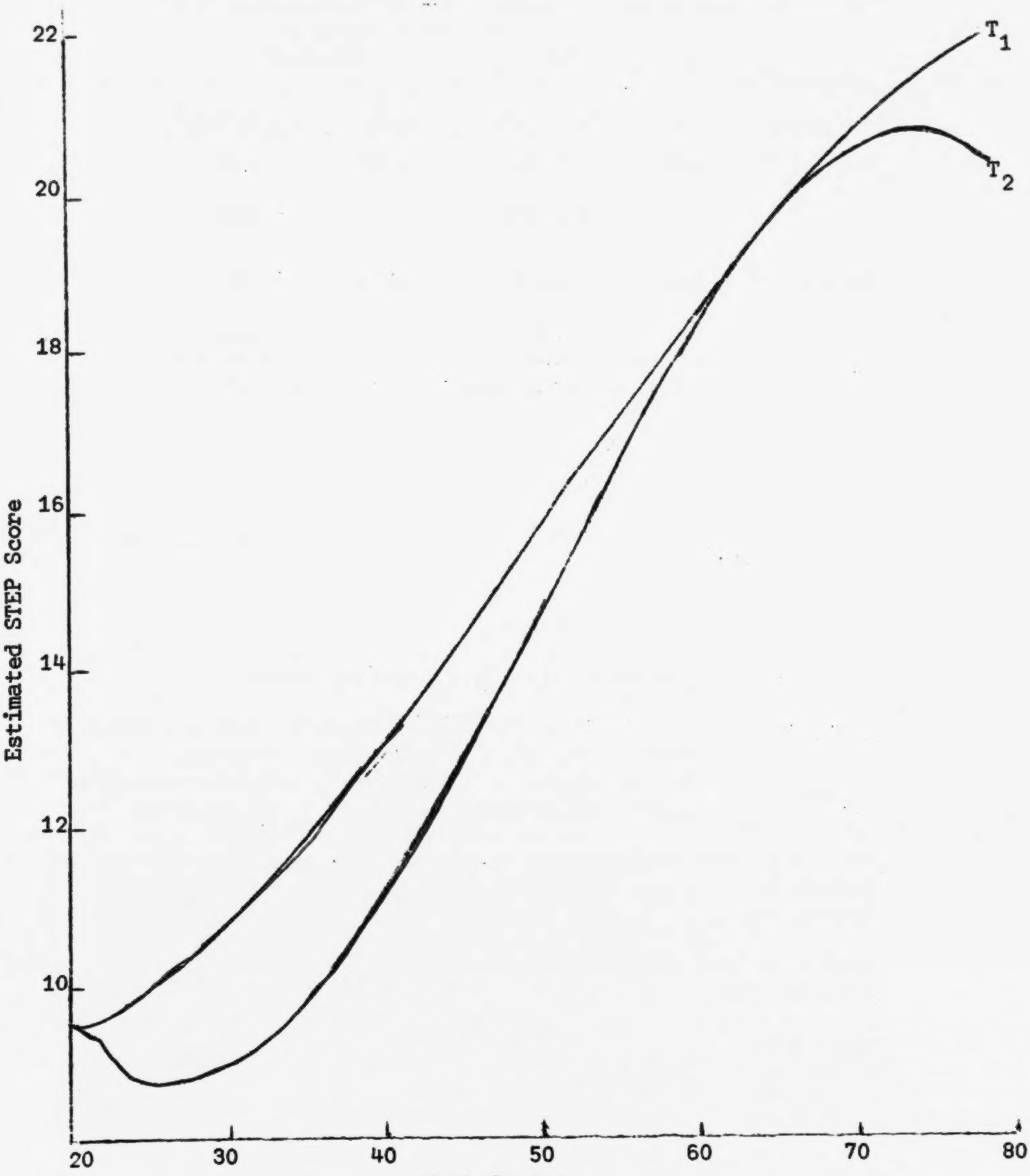


Figure 5. Regression of STEP Writing on Math for Study 1

Table 22

F-Ratios and Proportions of Variance (R^2's) Accounted for by Full and Reduced Models for STEP and Stanford Tests in Study 1

		Criterion Variable			
		STEP		Stanford	
Independent Variable		Full Model	$TX+TX^2+TX^3$	Full Model	$TX+TX^2+TX^3$
Math	R^2	.4886	.4530	.4368	.4309
	F		2.6900*		.4300
English	R^2	.6028	.6007	.4932	.4834
	F		.2000		.7500

*Significant at beyond .05 level, N= 124

Table 23

Item Selection Data for SI Ability Tests

Test	Total No. of Items	No. of Items Selected	Difficulty Range of Selected Items	Average Intercorrelation of Selected Items
Correlate Completion	40	19	.40-.60	.391
Word Classification	40	13	.38-.70	.054
Class Name Selection	30	11	.39-.71	.161

scores perform at about the same level on the STEP writing test under either treatment. The analysis section of the preceding chapter specified that three other models would be fitted and interactions tested for the STEP, Stanford and Structural Relationships tests. They involved using both Math and English in one model, using Math and the Structural Relationships pretest in one model, and using the factor scores computed for the SI ability measures and either Math or English in one model. All of these analyses were made for the data of study 1. None of them produced significant aptitude treatment interactions.

Analyses of Modified Ability Variables

The two part scores for each ability measure were intercorrelated over the entire group of subjects in study 2. The correlations were .80 for Correlate Completion, .47 for Word Classification, .66 for Seeing Trends, and .53 for Class Name Selection. Rather close agreement exists between these correlations and the reliabilities reported in Table 1 except for Word Classification, where the reliability estimate was .76. The reliabilities for the total tests, computed by the Spearman Brown formula, were .89 for Correlate Completion, .64 for Word Classification, .80 for Seeing Trends, and .69 for Class Name Selection.

Two conditions must be met before a symmetric bimodal distribution of test scores can be obtained. First, items of mid-range difficulty should exist and second, their average intercorrelation should be relatively high. Item difficulties and intercorrelations for each of the total tests were examined to determine whether bimodal distributions could be produced. Table 23 shows the results of these analyses for three of the tests. Seeing Trends was eliminated because of the extreme difficulty of most of the items in both parts of the test. The item analysis procedures indicated that only Correlate Completion was capable of being modified to have a bimodal distribution so that no further analyses were done on the other two tests. For Correlate Completion, the nineteen items with difficulties between .40 and .60 were used to form the bimodal version of the test.

In order to study the effect of modifying the Correlate Completion distribution on the production of ATI effects the following model was used to estimate the Aluminum Rewrite criterion scores:

$$\hat{Y} = a + b_1T + b_2X + b_3Z + (b_4TX + b_5TZ + b_6TZX)$$

In this model T is a dummy variable with values of +.5 and -.5 for T_1 and T_2 respectively, X is the pretest variable and Z is

Correlate Completion. A reduced model employing only the main effect variables was also fitted, and the difference between the squared multiple correlations of the two models was used as an indication of interaction effects. The two models were used with the bimodal form of the Correlate Completion test and with the part 1 and total scores. The results of these analyses are shown in Table 24.

Table 24

Squared Multiple Correlations for the Regression of the Aluminum Rewrite Criteria on Three Versions of Correlate Completion

Model	Criterion Measure								
	W/T Correlate Completion			W/C Correlate Completion			C/T Correlate Completion		
	Part 1	Bimodal	Total	Part 1	Bimodal	Total	Part 1	Bimodal	Total
Full	.536	.541	.539	.435	.448	.445	.429	.467	.475
Reduced	.511	.523	.517	.411	.420	.414	.406	.416	.416
Difference	.025	.018	.022	.024	.028	.034	.023	.051	.059

These analyses indicated that only in the case of the C/T criterion did the bimodal and total Correlate Completion scores produce greater ATI effects than the part 1 score. It should be pointed out that the means and standard deviations for part 1 and total versions of the test indicated that their distributions tended to be platykurtic. Table 25 contains these means and standard deviations. Revision of the total test to make it bimodal apparently resulted in pulling in the ends of the distribution rather than "hollowing out" its middle as would have been the case if its distribution had been normal.

Effectiveness of the Grammar Treatments

Although it was not a major purpose of this investigation to determine whether the study of grammar generally promotes student growth in syntactic maturity and language usage, it is worthwhile to make some inquiries in those directions. Table 26 shows that the largest difference between pre and posttest W/T means was .83 for T_1 in study 1. The next largest was .51 for T_2 in study 1. Hunt (1968) investigated the growth of syntactic

Table 25

Means and Standard Deviations for Three Versions of Correlate Completion

Correlate Completion	Mean	S. D.	No. of Items
Part 1	8.94	6.23	20
Bimodal	8.04	5.82	19
Total	16.25	11.01	40

Table 26

Mean Words per T-Unit (W/T) Scores For Subjects of Study 1, Study 2, and Control Groups in Time Sequence

	Time of Testing				
Source of Data	September		February		May
		Difference		Difference	
Study 1					
T_1	9.43	.83	10.26		
T_2	9.25	.51	9.76		
M	9.34	.67	10.01		
Study 2					
T_1			9.26	.11	9.37
T_2			9.49	.26	9.75
M			9.38	.18	9.56
Control			9.54	.18	9.72

maturity in cross sectional samples of students in grades four, six, eight, ten and twelve and found W/T means of 9.84 and 10.44, and 11.30 for grades eight, ten and twelve respectively. Thus between grades eight and ten a difference of .60 was found and between ten and twelve it was .86. It can be argued, therefore, that the differences found for the two treatment groups of study 1 probably represent accelerations of at least one year of growth in the W/T measure of syntactic maturity. Effects of testing and maturation seem to be minimal in accounting for the change since the pre-post difference for the control group was only .18 and the difference between the average of the pretest treatment means for the data of the two studies was .04. Similar results can be shown for the other two syntactic maturity measures, W/C and C/T. In study 1 only the T_1 means give any indication of growth of W/C but both treatments appear to show some growth in C/T means when compared with Hunt's (1968) norms (for W/C Hunt found means of 6.79, 7.35 and 7.85 at grade levels eight, ten, and twelve and for C/T means of 1.43, 1.42, and 1.44 at the same grades.) None of the three syntactic maturity variables show evidence of growth in the subjects of study 2. It is possible that the inability of many of the study 2 subjects to complete the treatments in the time allotted is the reason for the discrepancies between the two studies.

Caution must be used in accepting these results at face value since equal differences may not indicate the same growth for groups of subjects who differ in initial level of the variable under consideration. Hunt's eighth grade group had a higher W/T mean than the pretest means of the tenth graders of this study, and it may be that the .60 gain from eight to tenth grade represents a greater increment of growth than the .83 difference observed in T_1 means since the T_1 group started at a lower level. In addition, the finding of strong curvilinear relationships between pre and post W/T scores indicates that most of the growth shown in the treatment group occurred in subjects who were initially low on the scale.

V. Discussion

The first section of this chapter presents the objectives of the investigation in series and relates the results of the study to each of them. The second section points out the limitations of the study.

Objectives of the Investigation

Objective One

The first objective was to determine whether ATI effects on syntactic maturity and on knowledge of structural relationships in English occur after several months of instruction. The results of the analyses of the data of study 1 did show some evidence of ATI effects after a period of approximately five months of instruction, but they were of relatively small magnitudes. The interaction variables typically accounted for two or three percent of the variance of the criterion measures while the pretest and ability variables together typically accounted for much more, usually around fifty percent.

The classification procedure used to help interpret the nature of the interaction indicated that few subjects would profit more from the study of transformational grammar than from the traditional kind. This fact, as well as the finding that the traditional group showed a greater mean increase than the transformational group did, seems to be in disagreement with the results obtained by Mellon (1967). He found that instruction for nine months in transformational grammar in addition to sentence combining exercises produced a mean gain of 1.27 words per T-unit in seventh grade students, while instruction in traditional grammar resulted in a mean gain of .26. In addition, he failed to find a significant pretest by treatment interaction, although initially high students gained more under transformational grammar than students who were initially low (1.35 vs. 1.19), while low students in traditional grammar gained the most (.49 vs. .02). One reason for the discrepancy between the findings of the two investigations may lie in the amount and kind of experience in sentence combining given in the various treatments.

Mellon's subjects who received transformational grammar and sentence-combining exercises learned to transform seven kernel sentences into one complex statement by the seventh month of study. His traditional group, however, used textbooks which required the subjects to deal primarily with simple sentences throughout the entire period of instruction. In the present study the situation was reversed. English 3200 contains a considerable number of frames in which the student is asked to rewrite non-kernel sentences in a better way or to combine two kernel sentences into one. Modern English Sentence Structure only occasionally contains a frame which allows this kind of practice. Viewed in this light the findings of the two investigations appear to be consistent and to support Mellon's contention that sentence-combining exercises are more important in increasing syntactic maturity than the kind of grammar used in connection with them.

Ability measures from the structure of intellect model did not appear to be differently related to success on the criterion measures for the two treatments. It was originally expected that tests in the symbolic category might be more highly related to the criteria for subjects who studied Modern English Sentence Structure and that those in the semantic category would be more highly related to success in English 3200. This expectation, however, was not borne out since inspection of the regression equations involving each ability in turn failed to reveal different patterns of coefficients for tests of the two content categories. Similar patterns of coefficients were found for regression equations involving the two factor scores computed from the SI abilities and interpreted as symbolic and semantic content factors. These findings agree with those of Cronbach and Snow (1969) who suggested that general mental ability is the source of most ATI effects that have been found in the literature to the present time.

Analyses of the STEP and Stanford criterion measures showed a significant ATI effect for STEP when mathematics was used as the ability variable. The interaction was ordinal in that subjects of low and medium math ability did less well in transformational than traditional grammar, but subjects high in mathematics did equally well under either treatment. In a sense this finding partially confirms the expectation that success in transformational grammar is more dependent on symbolic abilities than is success in traditional grammar.

Objective Two

The second objective was to modify and refine the SI ability measures used as predictors in order to increase their differential validity in measuring ATI. Discussion of findings in relation to the first objective indicated that no evidence for differential validity of the SI tests could be found. Therefore, no attempt could be made to increase that kind of validity.

An attempt was made to determine whether bimodal distributions could be produced from the items of the SI ability tests which were retained and lengthened for study 2. It was expected that ATI effects would be enhanced if the crossover point of the treatment regression lines for predicting a criterion variable from an ability occurred where few subjects were located in the ability distribution. A more detailed explanation was given in the analysis section. Item difficulties and intercorrelations indicated that only one test, Correlate Completion, could be modified to produce a bimodal distribution. It was also determined that the original distribution was platykurtic so that bimodality was produced by pulling in the ends of the distribution rather than by "hollowing" out the middle.

Objective Three

The third objective was to replicate the study, if time permitted, to determine whether ATI effects are consistent and to determine whether the revised ability measures are better indicators of ATI than the unrevised ones. The classification procedure previously described demonstrated that the regression weights derived in study 1 could not be used for optimal placement of subjects in grammatical treatments in study 2.

The most obvious reason for the cross validation failure is that the weak ATI effects of study 1 were produced by idiosyncratic characteristics of the sample which were not representative of the population. Another reason for the cross validation failure could lie in the shorter duration of the treatments in study 2. The overall mean gains of both treatment groups in study 2 were comparable to the gains of the control group that was given both pretest and posttest within a period of two weeks, while the gains of both treatment groups in study 1 were considerably greater. Thus, it is possible that had the treatments of study 2 been continued for approximately two more months or until all subjects had finished the textbooks, ATI effects similar to those of study 1 would have occurred.

One effect that occurred in study 1 did cross validate in study 2. Subjects who were predicted to achieve better under treatment 1 had higher W/T and W/C means than unclassified subjects or those who were predicted to achieve better under treatment 2 regardless of the actual treatment they received. While these results do not indicate an ATI effect in themselves, they suggest that while either treatment is adequate for subjects classified as "belonging to treatment 1," perhaps a third treatment such as Mellon's sentence-combining exercises should be sought for unclassified subjects and those who "belong to treatment 2." On the W/T criterion, study 1 subjects who "belong to treatment 1" were those who were above the mean on Correlate Completion. For W/C the subjects who were classified as "belonging to treatment 1" were above the mean on Class Name Selection.

The bimodal and total versions of the Correlate Completion test appeared to be clearly superior to part 1 of the test in producing ATI effects in only the C/T criterion of Aluminum Rewrite. For the other two criteria, W/C and W/T, the magnitudes of the ATI effects were about the same for all three versions. For all of the Aluminum Rewrite criteria, however, the proportions of variance explained by the independent variables were consistently, though not greatly, higher for the revisions of Correlate Completion than for part 1 only.

It is likely that the bimodal version failed to show superiority over the total version because the total test distribution was platykurtic and did not place a great many subjects on or near the crossover line of the regression planes. It is of interest to note that even though the bimodal version was only half as long as the total, it was almost as effective as the total test in all respects. Thus if Correlate Completion were to be used in future ATI studies, the bimodal test could be used in place of the total since it could be administered in half the time that would be required for the total test.

Objective Four

The fourth objective, if ATI effects were discovered, was to conduct utility studies to determine whether the increased cost of using two kinds of instructional materials outweighed the increments in learning that they produced. Since the ATI effects found in study 1 could not be cross validated on the data of study 2 and since few subjects were identified who would profit more from treatment two than treatment one, it was concluded that there was no basis for recommending that the procedures used in this study be put into everyday educational practice. Therefore, there was no reason to pursue the fourth objective.

Limitations of the Study

This section is devoted to problems encountered in the investigation that might be at least partly responsible for the failure to find ATI effects of sufficient stability and magnitude to warrant practical application of them. These problems center around the treatments, the criterion tests, and the analytic models used in the investigation.

Treatments

One of the treatments, Modern English Sentence Structure, appeared to be much too difficult for most of the subjects to whom it was assigned. The absence of adequate unit tests which the subjects could use for self-evaluation probably contributed to the ineffectiveness of this treatment. It should be pointed out that this problem might not have been so severe if the study had been conducted in another setting in which different subjects were used. The subjects that were used might well be atypical of tenth grade students generally since their syntactic maturity mean scores were lower than those reported by Hunt (1968) for eighth grade students. English 3200 did not appear to be too difficult for any but the very slowest students.

Neither treatment was well liked by the students or the teachers. The programed format of both textbooks seemed to be responsible for this defect. It was suggested to the teachers that they allow small groups of subjects who were at about the same place in the same textbook to work together and that they offer free time as incentives to master the material. However, crowded classroom conditions prevented the teachers from using these or other techniques for motivating the subjects.

Criterion Tests

The strong curvilinear relationships between the pretest and posttest Aluminum Rewrite variables were unanticipated at the onset of the study. Other researchers (Mellon, 1967; Blount, Frederick, and Johnson, 1969) who have used words per T-unit (W/T), words per clause (W/C), and clauses per T-unit (C/T) as both pretest and criterion variables failed to investigate the possibility of non-linear relationships between them. These investigators, however, developed the three ratios from samples of free writing and it is possible that only linear relationships exist between the variables under that condition. The purpose of the Aluminum Rewrite test is to approximate the measures of syntactic maturity that are obtained from free writing but that are more time-consuming and expensive to obtain. If it were found to give widely discrepant results because of the curvilinear relationships between pre and post measures its scoring procedures would need to be drastically revised or it should probably be abandoned in favor of the use of free writing to obtain the criteria.

Analytic Models

It is uncertain that the models used in analyzing the data were optimal for discovering ATI effects. For example, more complex models could have been constructed by including variables representing the first order interaction of ability and pretest and by including all or several ability measures and their interactions with each other and the pretest and treatment. On the other hand, simpler models could have been made by deleting non-linear terms and by using only first order interaction variables. Probably the models should have been made simpler rather than more complex in order to enhance the likelihood of greater stability of the regression coefficients. In that case, however, it would have been necessary to make some kind of transformation of the data in order to reduce or eliminate the non-linear relationships between pretest and criterion measures.

REFERENCES

Bateman, D. R., and Zidonis, F. J. The Effect of a Study of Transformational Grammar on the Writing of Ninth and Tenth Graders. NCTE Research Report No. 6; Champaign, Illinois: National Council of Teachers of English, 1966.

Bennett, R. A. A Comparison of the Effects of Programed Instruction and Lecture-Textbook Instruction on English Language Learning of Eleventh Grade Students. Unpublished doctoral dissertation, Florida State University, Tallahassee, 1964.

Blount, N. C., Fredrick, W. C., and Johnson, S. L. The effect of a study of grammar on the writing of eighth grade students. Technical Report from the Research and Development Center for Cognitive Learning. University of Wisconsin. No. 69. 1968.

Blumenthal, J. C. English 3200: A Programmed Course in Grammar and Usage. New York: Harcourt Brace and World, Inc. 1962.

Carroll, J. B. Research on teaching foreign languages. In Gage, N. L. (Ed.) Handbook of Research on Teaching. Chicago: Rand McNally and Co. 1963.

Cronbach, L. J., and Snow, R. E. Individual Differences in Learning Ability as a Function of Instructional Variables. Contract No. U.S. OEC 4-6-061269-1217, United States Office of Education. Stanford, California. Stanford University. 1969

Davis, J. B., Jr. An Investigation of the Interaction of Certain Instructional Strategies with the Structure of Basic Mental Abilities in the Learning of Some Mathematical Operations. Unpublished doctoral dissertation, Florida State University, Tallahassee, 1967.

Frederick, W. C., Blount, N. S., and Johnson, S. L. A Comparison of Verbal Statement, Symbolic Notation and Figural Representation of Grammar Concepts. Technical Report from the Research and Development Center for Cognitive Learning. University of Wisconsin No. 64. 1968.

Guilford, J. P., and Hoepfner, R. Structure of Intellect Factors and Their Tests. Reports from the Psychological Laboratory of the University of Southern California. No. 36. 1966.

Guilford, J. P., Kettner, N. W., and Christensen, P. R. A Factor-Analytic Study Across the Domains of Reasoning, Creativity, and Evaluation. I: Hypotheses and Description of Tests. Reports from the Psychological Laboratory of the University of Southern California. No. 11, July, 1954.

Guilford, J. P., Merrifield, P. R., Christensen, P. R., and Frick, J. W. An Investigation of Symbolic Factors of Cognition and Convergent Production. Reports from the Psychological Laboratory of the University of Southern California. No. 23, April, 1960.

Hoepfner, R., Guilford, J. P., and Merrifield, P. R. A Factor-Analysis of the Symbolic-Evaluation Abilities. Reports from the Psychological Laboratory of the University of Southern California. No. 33. December, 1964.

Hoffman, W. E. A Study of the effectiveness of programmed instruction in English grammar and usage in the high school English classroom. Dissertation Abstracts, 29, 1040A, 1968.

Hunt, K. W. Differences in Grammatical Structures Written at Three Grade Levels. Cooperative Research Project No. 1998, United States Office of Education. Tallahassee, Florida. Florida State University. 1964.

Hunt, K. W. Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3; Champaign, Illinois: National Council of Teachers of English, 1965.

Hunt, K. W. Sentence Structures Used by Superior Students in Grades Four and Twelve and by Superior Adults. Cooperative Research Project No. 5-0313, United States Office of Education. Tallahassee, Florida. Florida State University. 1966.

Hunt, K. W., and others. An Instrument to Measure Syntactic Maturity. Preliminary Version. Tallahassee, Florida. Florida State University, 1968.

Kahler, A. D., Jr. The effects of programed grammar and journal writing on students writing ability: An exploratory study. Dissertation Abstracts, 27, 74-A, 1966.

Kropp, R. P., Nelson, W. H., and King, F. J. Identification and Definition of Subject-Matter Content Variables Related to Human Aptitudes. Cooperative Research Project No. 2914, United States Office of Education. Tallahassee, Florida. Florida State University. 1967.

Mellon, J. C. Transformational Sentence-Combining: A Method for Enhancing the Development of Syntactic Fluency in English Composition. Cooperative Research Project No. 5-8418, United States Office of Education. Cambridge, Massachusetts. Harvard University. 1967.

Munday, R. G. The effects of the use of English 3200, a programed textbook, on achievement in English grammar at the twelfth-grade level in a large metropolitan high school. Dissertation Abstracts, 26, 3928, 1966.

Nihira, K., Guilford, J. P., Hoepfner, R., and Merrifield, P. R. A Factor Analysis of the Semantic-Evaluation Abilities. Reports from the Psychological Laboratory of the University of Southern California. No. 32. December, 1964.

O'Donnell, R. C. The Correlation of Awareness of Structural Relationships in English and Ability in Written Composition. Cooperative Research Project No. 1524, United States Office of Education. Mount Olive, North Carolina. Mount Olive College. 1963.

O'Donnell, R. C., Griffin, W. J., and Norris, R. C. Syntax of Kindergarten and Elementary School Children: A Transformational Analysis. NCTE Research Report No. 8; Champaign, Illinois: National Council of Teachers of English, 1967.

Reed, J. E. and Hayman, J. L., Jr. An experiment involving use of English 2600, an automated instruction text. The Journal of Educational Research, 1962, 55, 476-484.

Rogovin, Syrell. Modern English Sentence Structure. Syracuse: The L. W. Singer Company. 1964.